

1. Unit 1. The Chemistry of Life

1. The Study of Life

1. [Introduction](#)
2. [The Science of Biology](#)
3. [Themes and Concepts of Biology](#)

2. The Chemical Foundation of Life

1. [Introduction](#)
2. [Atoms, Isotopes, Ions, and Molecules: The Building Blocks](#)
3. [Water](#)
4. [Carbon](#)

3. Biological Macromolecules

1. [Introduction](#)
2. [Synthesis of Biological Macromolecules](#)
3. [Carbohydrates](#)
4. [Lipids](#)
5. [Proteins](#)
6. [Nucleic Acids](#)

2. Unit 2. The Cell

1. Cell Structure

1. [Introduction](#)
2. [Studying Cells](#)
3. [Prokaryotic Cells](#)
4. [Eukaryotic Cells](#)
5. [The Endomembrane System and Proteins](#)
6. [The Cytoskeleton](#)
7. [Connections between Cells and Cellular Activities](#)

2. Structure and Function of Plasma Membranes

1. [Introduction](#)
2. [Components and Structure](#)
3. [Passive Transport](#)

4. [Active Transport](#)
 5. [Bulk Transport](#)
3. Metabolism
 1. [Introduction](#)
 2. [Energy and Metabolism](#)
 3. [The Laws of Thermodynamics](#)
 4. [Potential, Kinetic, Free, and Activation Energy](#)
 5. [ATP: Adenosine Triphosphate](#)
 6. [Enzymes](#)
3. Unit 3. Genetics
 1. Meiosis and Sexual Reproduction
 1. [Introduction](#)
 2. [The Process of Meiosis](#)
 3. [Sexual Reproduction](#)
 2. Mendel's Experiments and Heredity
 1. [Introduction](#)
 2. [Mendel's Experiments and the Laws of Probability](#)
 3. [Characteristics and Traits](#)
 4. [Laws of Inheritance](#)
 3. Modern Understandings of Inheritance
 1. [Introduction](#)
 2. [Chromosomal Theory and Genetic Linkage](#)
 3. [Genetic Linkages and Distances](#)
 4. [Chromosomal Basis of Inherited Disorders](#)

Introduction

class="introduction"

This NASA image is a composite of several satellite-based views of Earth. To make the whole-Earth image, NASA scientists combine observations of different parts of the planet. (credit: NASA/GSFC/NOAA/USGS)

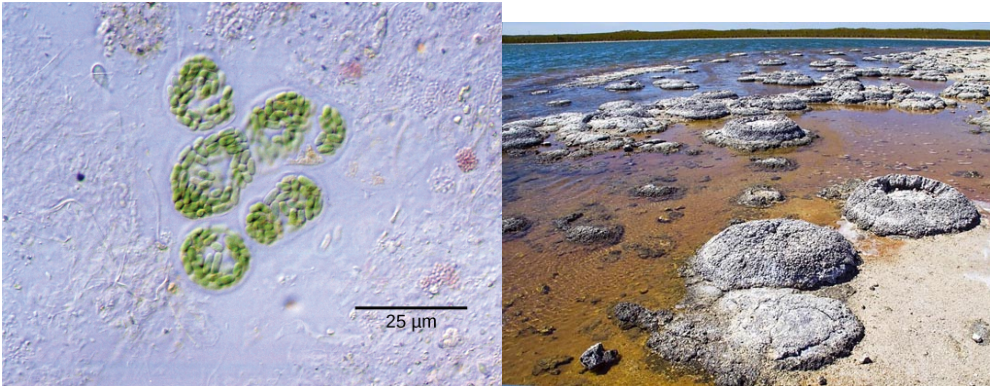


Viewed from space, Earth offers no clues about the diversity of life forms that reside there. The first forms of life on Earth are thought to have been microorganisms that existed for billions of years in the ocean before plants and animals appeared. The mammals, birds, and flowers so familiar to us are all relatively recent, originating 130 to 200 million years ago. Humans have inhabited this planet for only the last 2.5 million years, and only in the last 200,000 years have humans started looking like we do today.

The Science of Biology

By the end of this section, you will be able to:

- Identify the shared characteristics of the natural sciences
- Summarize the steps of the scientific method
- Compare inductive reasoning with deductive reasoning
- Describe the goals of basic science and applied science



Formerly called blue-green algae, these (a) cyanobacteria, shown here at 300x magnification under a light microscope, are some of Earth's oldest life forms.

These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters. (credit a: modification of work by NASA; credit b: modification of work by Ruth Ellison; scale-bar data from Matt Russell)

What is biology? In simple terms, **biology** is the study of living organisms and their interactions with one another and their environments. This is a very broad definition because the scope of biology is vast. Biologists may study anything from the microscopic or submicroscopic view of a cell to ecosystems and the whole living planet ([\[link\]](#)). Listening to the daily news, you will quickly realize how many aspects of biology are discussed every day. For example, recent news topics include *Escherichia coli* ([\[link\]](#)) outbreaks in spinach and *Salmonella* contamination in peanut butter. Other subjects include efforts toward finding a cure for AIDS, Alzheimer's

disease, and cancer. On a global scale, many researchers are committed to finding ways to protect the planet, solve environmental issues, and reduce the effects of climate change. All of these diverse endeavors are related to different facets of the discipline of biology.



Escherichia coli (*E. coli*) bacteria, seen in this scanning electron micrograph, are normal residents of our digestive tracts that aid in the absorption of vitamin K and other nutrients. However, virulent strains are sometimes responsible for disease outbreaks. (credit: Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU)

The Process of Science

Biology is a science, but what exactly is science? What does the study of biology share with other scientific disciplines? **Science** (from the Latin *scientia*, meaning “knowledge”) can be defined as knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method. It becomes clear from this definition

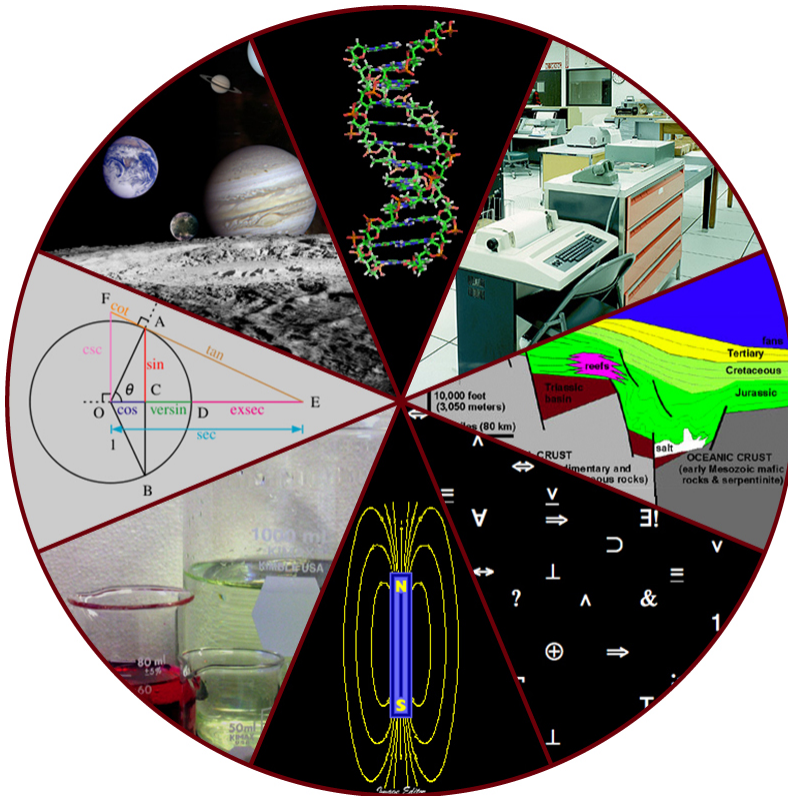
that the application of the scientific method plays a major role in science. The **scientific method** is a method of research with defined steps that include experiments and careful observation.

The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses by means of repeatable experiments. A **hypothesis** is a suggested explanation for an event, which can be tested. Although using the scientific method is inherent to science, it is inadequate in determining what science is. This is because it is relatively easy to apply the scientific method to disciplines such as physics and chemistry, but when it comes to disciplines like archaeology, psychology, and geology, the scientific method becomes less applicable as it becomes more difficult to repeat experiments.

These areas of study are still sciences, however. Consider archeology—even though one cannot perform repeatable experiments, hypotheses may still be supported. For instance, an archeologist can hypothesize that an ancient culture existed based on finding a piece of pottery. Further hypotheses could be made about various characteristics of this culture, and these hypotheses may be found to be correct or false through continued support or contradictions from other findings. A hypothesis may become a verified theory. A **theory** is a tested and confirmed explanation for observations or phenomena. Science may be better defined as fields of study that attempt to comprehend the nature of the universe.

Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A planetarium? Gems and minerals? Or, maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics ([\[link\]](#)). However, those fields of science related to the physical world and its phenomena and processes are considered **natural sciences**. Thus, a museum of natural sciences might contain any of the items listed above.



The diversity of scientific fields includes astronomy, biology, computer science, geology, logic, physics, chemistry, mathematics, and many other fields. (credit: “Image Editor”/Flickr)

There is no complete agreement when it comes to defining what the natural sciences include, however. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into **life sciences**, which study living things and include biology, and **physical sciences**, which study nonliving matter and include astronomy, geology, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on both life and physical sciences and are interdisciplinary. Natural sciences are sometimes referred to as “hard science” because they rely on the use of quantitative data; social sciences that study society and human behavior are more likely to use qualitative assessments to drive investigations and findings.

Not surprisingly, the natural science of biology has many branches or subdisciplines. Cell biologists study cell structure and function, while biologists who study anatomy investigate the structure of an entire organism. Those biologists studying physiology, however, focus on the internal functioning of an organism. Some areas of biology focus on only particular types of living things. For example, botanists explore plants, while zoologists specialize in animals.

Scientific Reasoning

One thing is common to all forms of science: an ultimate goal “to know.” Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. To do this, they use two methods of logical thinking: inductive reasoning and deductive reasoning.

Inductive reasoning is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in descriptive science. A life scientist such as a biologist makes observations and records them. These data can be qualitative or quantitative, and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Brain studies provide an example. In this type of research, many live brains are observed while people are doing a specific activity, such as viewing images of food. The part of the brain that “lights up” during this activity is then predicted to be the part controlling the response to the selected stimulus, in this case, images of food. The “lighting up” of the various areas of the brain is caused by excess absorption of radioactive sugar derivatives by active areas of the brain. The resultant increase in radioactivity is observed by a scanner. Then, researchers can stimulate that part of the brain to see if similar responses result.

Deductive reasoning or deduction is the type of logic used in hypothesis-based science. In deductive reason, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. **Deductive reasoning** is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. Studies in climate change can illustrate this type of reasoning. For example, scientists may predict that if the climate becomes warmer in a particular region, then the distribution of plants and animals should change. These predictions have been made and tested, and many such changes have been found, such as the modification of arable areas for agriculture, with change based on temperature averages.

Both types of logical thinking are related to the two main pathways of scientific study: descriptive science and hypothesis-based science. **Descriptive (or discovery) science**, which is usually inductive, aims to observe, explore, and discover, while **hypothesis-based science**, which is usually deductive, begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, and most scientific endeavors combine both approaches. The fuzzy boundary becomes apparent when thinking about how easily observation can lead to specific questions. For example, a gentleman in the 1940s observed that the burr seeds that stuck to his clothes and his dog's fur had a tiny hook structure. On closer inspection, he discovered that the burrs' gripping device was more reliable than a zipper. He eventually developed a company and produced the hook-and-loop fastener popularly known today as Velcro. Descriptive science and hypothesis-based science are in continuous dialogue.

The Scientific Method

Biologists study the living world by posing questions about it and seeking science-based responses. This approach is common to other sciences as well and is often referred to as the scientific method. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626) ([\[link\]](#)), who set up inductive methods for scientific inquiry. The scientific method is not exclusively used by

biologists but can be applied to almost all fields of study as a logical, rational problem-solving method.



Sir Francis Bacon
(1561–1626) is
credited with being
the first to define
the scientific
method. (credit:
Paul van Somer)

The scientific process typically starts with an observation (often a problem to be solved) that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

Proposing a Hypothesis

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, “The classroom is warm because no one turned on the air conditioning.” But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be, “The classroom is warm because there is a power failure, and so the air conditioning doesn’t work.”

Once a hypothesis has been selected, the student can make a prediction. A prediction is similar to a hypothesis but it typically has the format “If . . . then” For example, the prediction for the first hypothesis might be, “*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm.”

Testing a Hypothesis

A valid hypothesis must be testable. It should also be **falsifiable**, meaning that it can be disproven by experimental results. Importantly, science does not claim to “prove” anything because scientific understandings are always subject to modification with further information. This step—openness to disproving ideas—is what distinguishes sciences from non-sciences. The presence of the supernatural, for instance, is neither testable nor falsifiable. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. Each experiment will have one or more variables and one or more controls. A **variable** is any part of the experiment that can vary or change during the experiment. The **control group** contains every feature of the experimental group except it is not given the manipulation that is hypothesized about. Therefore, if the results of the experimental group differ from the control group, the difference must be due to the hypothesized manipulation, rather than some outside factor. Look for the variables and controls in the examples that follow. To test the first hypothesis, the student would find out if the air conditioning is on. If the air conditioning is turned on but does not work, there should be another reason, and this hypothesis should be rejected. To

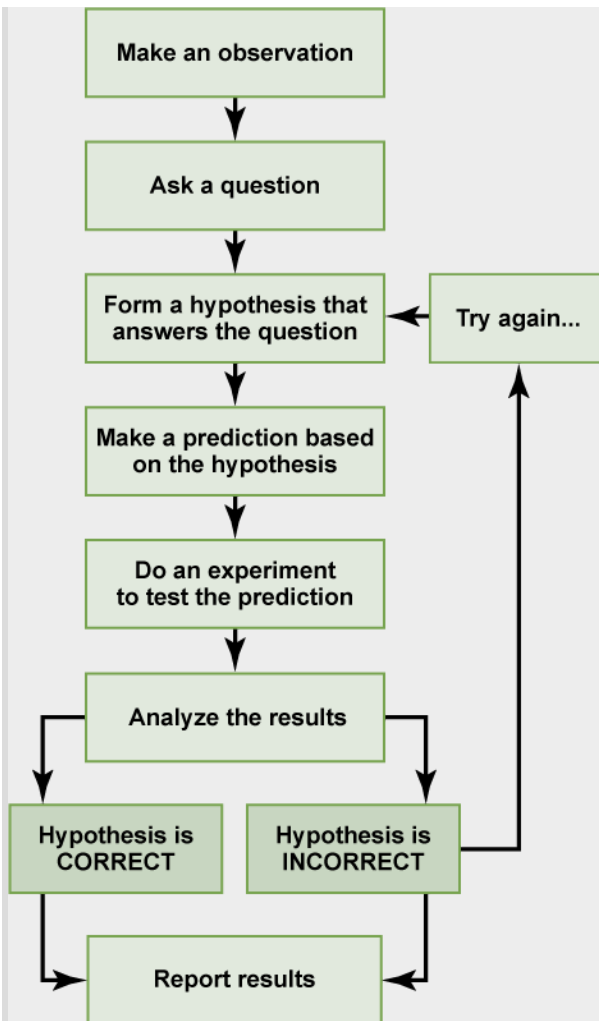
test the second hypothesis, the student could check if the lights in the classroom are functional. If so, there is no power failure and this hypothesis should be rejected. Each hypothesis should be tested by carrying out appropriate experiments. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid ([\[link\]](#)). Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.

While this “warm classroom” example is based on observational results, other hypotheses and experiments might have clearer controls. For instance, a student might attend class on Monday and realize she had difficulty concentrating on the lecture. One observation to explain this occurrence might be, “When I eat breakfast before class, I am better able to pay attention.” The student could then design an experiment with a control to test this hypothesis.

In hypothesis-based science, specific results are predicted from a general premise. This type of reasoning is called deductive reasoning: deduction proceeds from the general to the particular. But the reverse of the process is also possible: sometimes, scientists reach a general conclusion from a number of specific observations. This type of reasoning is called inductive reasoning, and it proceeds from the particular to the general. Inductive and deductive reasoning are often used in tandem to advance scientific knowledge ([\[link\]](#)). In recent years a new approach of testing hypotheses has developed as a result of an exponential growth of data deposited in various databases. Using computer algorithms and statistical analyses of data in databases, a new field of so-called “data research” (also referred to as “in silico” research) provides new methods of data analyses and their interpretation. This will increase the demand for specialists in both biology and computer science, a promising career opportunity.

Note:

Art Connection



The scientific method consists of a series of well-defined steps. If a hypothesis is not supported by experimental data, a new hypothesis can be proposed.

In the example below, the scientific method is used to solve an everyday problem. Order the scientific method steps (numbered items) with the process of solving the everyday problem (lettered items). Based on the results of the experiment, is the hypothesis correct? If it is incorrect, propose some alternative hypotheses.

1. Observation

2. Question
3. Hypothesis (answer)
4. Prediction
5. Experiment
6. Result

- a. There is something wrong with the electrical outlet.
- b. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
- c. My toaster doesn't toast my bread.
- d. I plug my coffee maker into the outlet.
- e. My coffeemaker works.
- f. Why doesn't my toaster work?

Note:

Art Connection

Two Types of Reasoning	
Inductive reasoning: from a number of observations, a general conclusion is drawn.	Deductive reasoning: from a general premise, specific results are predicted.
Observations	General premise
<ul style="list-style-type: none"> Members of a species are not all the same. Individuals compete for resources. Species are generally adapted to their environment. 	Individuals most adapted to their environment are more likely to survive and pass their traits on to the next generation.
↓	↓
Conclusion	Predicted results
Individuals most adapted to their environment are more likely to survive and pass their traits to the next generation.	If the average temperature in an ecosystem increases due to climate change, individuals better adapted to warmer temperatures will outcompete those that are not.

Scientists use two types of reasoning, inductive and deductive reasoning, to advance scientific knowledge. As is the case in this example, the conclusion from inductive reasoning can often become the premise for inductive reasoning.

Decide if each of the following is an example of inductive or deductive reasoning.

1. All flying birds and insects have wings. Birds and insects flap their wings as they move through the air. Therefore, wings enable flight.
2. Insects generally survive mild winters better than harsh ones. Therefore, insect pests will become more problematic if global temperatures increase.
3. Chromosomes, the carriers of DNA, separate into daughter cells during cell division. Therefore, DNA is the genetic material.
4. Animals as diverse as humans, insects, and wolves all exhibit social behavior. Therefore, social behavior must have an evolutionary advantage.

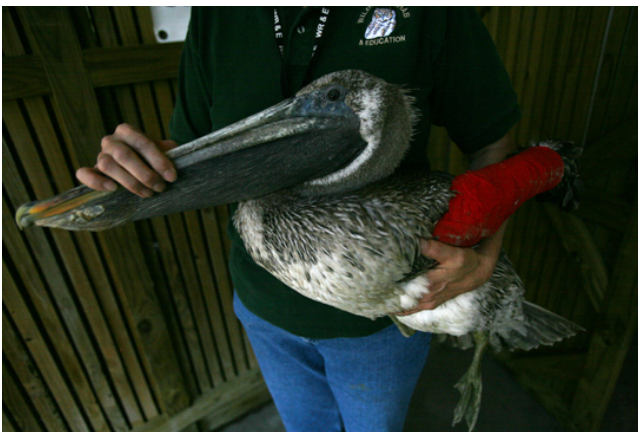
The scientific method may seem too rigid and structured. It is important to keep in mind that, although scientists often follow this sequence, there is flexibility. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests. Notice, too, that the scientific method can be applied to solving problems that aren't necessarily scientific in nature.

Two Types of Science: Basic Science and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or to bettering our lives? This question focuses on the differences between two types of science: basic science and applied science.

Basic science or “pure” science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge’s sake, though this does not mean that, in the end, it may not result in a practical application.

In contrast, **applied science** or “technology,” aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster ([\[link\]](#)). In applied science, the problem is usually defined for the researcher.



After Hurricane Ike struck the Gulf Coast in 2008, the U.S. Fish and Wildlife Service rescued this

brown pelican. Thanks to applied science, scientists knew how to rehabilitate the bird. (credit: FEMA)

Some individuals may perceive applied science as “useful” and basic science as “useless.” A question these people might pose to a scientist advocating knowledge acquisition would be, “What for?” A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the wide knowledge foundation generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, DNA makes new copies of itself, shortly before a cell divides. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits and the exact location of each gene. (The gene is the basic unit of heredity; an individual’s complete collection of genes is his or her genome.) Other less complex organisms have also been studied as part of this project in

order to gain a better understanding of human chromosomes. The Human Genome Project ([link](#)) relied on basic research carried out with simple organisms and, later, with the human genome. An important end goal eventually became using the data for applied research, seeking cures and early diagnoses for genetically related diseases.



The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project, which sequenced the entire human genome, was completed in 2003. (credit: the U.S. Department of Energy Genome Programs (<http://genomics.energy.gov>))

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by **serendipity**, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew on the dish, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings in order for other researchers to expand and build upon their discoveries. Collaboration with other scientists—when planning, conducting, and analyzing results—are all important for scientific research. For this reason, important aspects of a scientist's work are communicating with peers and disseminating results to peers. Scientists can share results by presenting them at a scientific meeting or conference, but this approach can reach only the select few who are present. Instead, most scientists present their results in peer-reviewed manuscripts that are published in scientific journals. **Peer-reviewed manuscripts** are scientific papers that are reviewed by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

A scientific paper is very different from creative writing. Although creativity is required to design experiments, there are fixed guidelines when it comes to presenting scientific results. First, scientific writing must be

brief, concise, and accurate. A scientific paper needs to be succinct but detailed enough to allow peers to reproduce the experiments.

The scientific paper consists of several specific sections—introduction, materials and methods, results, and discussion. This structure is sometimes called the “IMRaD” format. There are usually acknowledgment and reference sections as well as an **abstract** (a concise summary) at the beginning of the paper. There might be additional sections depending on the type of paper and the journal where it will be published; for example, some review papers require an outline.

The **introduction** starts with brief, but broad, background information about what is known in the field. A good introduction also gives the rationale of the work; it justifies the work carried out and also briefly mentions the end of the paper, where the hypothesis or research question driving the research will be presented. The introduction refers to the published scientific work of others and therefore requires citations following the style of the journal. Using the work or ideas of others without proper citation is considered **plagiarism**.

The **materials and methods** section includes a complete and accurate description of the substances used, and the method and techniques used by the researchers to gather data. The description should be thorough enough to allow another researcher to repeat the experiment and obtain similar results, but it does not have to be verbose. This section will also include information on how measurements were made and what types of calculations and statistical analyses were used to examine raw data. Although the materials and methods section gives an accurate description of the experiments, it does not discuss them.

Some journals require a results section followed by a discussion section, but it is more common to combine both. If the journal does not allow the combination of both sections, the **results** section simply narrates the findings without any further interpretation. The results are presented by means of tables or graphs, but no duplicate information should be presented. In the **discussion** section, the researcher will interpret the results, describe how variables may be related, and attempt to explain the observations. It is indispensable to conduct an extensive literature search to

put the results in the context of previously published scientific research. Therefore, proper citations are included in this section as well.

Finally, the **conclusion** section summarizes the importance of the experimental findings. While the scientific paper almost certainly answered one or more scientific questions that were stated, any good research should lead to more questions. Therefore, a well-done scientific paper leaves doors open for the researcher and others to continue and expand on the findings.

Review articles do not follow the IMRAD format because they do not present original scientific findings, or primary literature; instead, they summarize and comment on findings that were published as primary literature and typically include extensive reference sections.

Section Summary

Biology is the science that studies living organisms and their interactions with one another and their environments. Science attempts to describe and understand the nature of the universe in whole or in part by rational means. Science has many fields; those fields related to the physical world and its phenomena are considered natural sciences.

Science can be basic or applied. The main goal of basic science is to expand knowledge without any expectation of short-term practical application of that knowledge. The primary goal of applied research, however, is to solve practical problems.

Two types of logical reasoning are used in science. Inductive reasoning uses particular results to produce general scientific principles. Deductive reasoning is a form of logical thinking that predicts results by applying general principles. The common thread throughout scientific research is the use of the scientific method, a step-based process that consists of making observations, defining a problem, posing hypotheses, testing these hypotheses, and drawing one or more conclusions. The testing uses proper controls. Scientists present their results in peer-reviewed scientific papers published in scientific journals. A scientific research paper consists of several well-defined sections: introduction, materials and methods, results,

and, finally, a concluding discussion. Review papers summarize the research done in a particular field over a period of time.

Art Connections

Exercise:

Problem:

[\[link\]](#) In the example below, the scientific method is used to solve an everyday problem. Order the scientific method steps (numbered items) with the process of solving the everyday problem (lettered items). Based on the results of the experiment, is the hypothesis correct? If it is incorrect, propose some alternative hypotheses.

1. Observation
 2. Question
 3. Hypothesis (answer)
 4. Prediction
 5. Experiment
 6. Result
-
- a. There is something wrong with the electrical outlet.
 - b. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
 - c. My toaster doesn't toast my bread.
 - d. I plug my coffee maker into the outlet.
 - e. My coffeemaker works.
 - f. Why doesn't my toaster work?

Solution:

[\[link\]](#) 1: C; 2: F; 3: A; 4: B; 5: D; 6: E. The original hypothesis is incorrect, as the coffeemaker works when plugged into the outlet. Alternative hypotheses include that the toaster might be broken or that the toaster wasn't turned on.

Exercise:

Problem:

[\[link\]](#) Decide if each of the following is an example of inductive or deductive reasoning.

1. All flying birds and insects have wings. Birds and insects flap their wings as they move through the air. Therefore, wings enable flight.
2. Insects generally survive mild winters better than harsh ones. Therefore, insect pests will become more problematic if global temperatures increase.
3. Chromosomes, the carriers of DNA, separate into daughter cells during cell division. Therefore, DNA is the genetic material.
4. Animals as diverse as humans, insects, and wolves all exhibit social behavior. Therefore, social behavior must have an evolutionary advantage.

Solution:

[\[link\]](#) 1: inductive; 2: deductive; 3: deductive; 4: inductive.

Review Questions**Exercise:**

Problem: The first forms of life on Earth were _____.

- a. plants
- b. microorganisms
- c. birds
- d. dinosaurs

Solution:

B

Exercise:

Problem:

A suggested and testable explanation for an event is called a _____.

- a. hypothesis
- b. variable
- c. theory
- d. control

Solution:

A

Exercise:

Problem:

Which of the following sciences is not considered a natural science?

- a. biology
- b. astronomy
- c. physics
- d. computer science

Solution:

D

Exercise:

Problem:

The type of logical thinking that uses related observations to arrive at a general conclusion is called _____.

- a. deductive reasoning
 - b. the scientific method
 - c. hypothesis-based science
 - d. inductive reasoning
-

Solution:

D

Exercise:

Problem:

The process of _____ helps to ensure that a scientist's research is original, significant, logical, and thorough.

- a. publication
 - b. public speaking
 - c. peer review
 - d. the scientific method
-

Solution:

C

Exercise:

Problem:

A person notices that her houseplants that are regularly exposed to music seem to grow more quickly than those in rooms with no music. As a result, she determines that plants grow better when exposed to music. This example most closely resembles which type of reasoning?

- a. inductive reasoning
- b. deductive reasoning
- c. neither, because no hypothesis was made
- d. both inductive and deductive reasoning

Solution:

A

Free Response**Exercise:****Problem:**

Although the scientific method is used by most of the sciences, it can also be applied to everyday situations. Think about a problem that you may have at home, at school, or with your car, and apply the scientific method to solve it.

Solution:

Answers will vary, but should apply the steps of the scientific method. One possibility could be a car which doesn't start. The hypothesis could be that the car doesn't start because the battery is dead. The experiment would be to change the battery or to charge the battery and then check whether the car starts or not. If it starts, the problem was due to the battery, and the hypothesis is accepted.

Exercise:**Problem:**

Give an example of how applied science has had a direct effect on your daily life.

Solution:

Answers will vary. One example of how applied science has had a direct effect on daily life is the presence of vaccines. Vaccines to prevent diseases such as polio, measles, tetanus, and even influenza affect daily life by contributing to individual and societal health.

Exercise:**Problem:**

Name two topics that are likely to be studied by biologists, and two areas of scientific study that would fall outside the realm of biology.

Solution:

Answers will vary. Topics that fall inside the area of biological study include how diseases affect human bodies, how pollution impacts a species' habitat, and how plants respond to their environments. Topics that fall outside of biology (the "study of life") include how metamorphic rock is formed and how planetary orbits function.

Exercise:**Problem:**

Thinking about the topic of cancer, write a basic science question and an applied science question that a researcher interested in this topic might ask

Solution:

Answers will vary. Basic science: What evolutionary purpose might cancer serve? Applied science: What strategies might be found to prevent cancer from reproducing at the cellular level?

Glossary

abstract

opening section of a scientific paper that summarizes the research and conclusions

applied science

form of science that aims to solve real-world problems

basic science

science that seeks to expand knowledge and understanding regardless of the short-term application of that knowledge

biology

the study of living organisms and their interactions with one another and their environments

conclusion

section of a scientific paper that summarizes the importance of the experimental findings

control

part of an experiment that does not change during the experiment

deductive reasoning

form of logical thinking that uses a general inclusive statement to forecast specific results

descriptive science

(also, discovery science) form of science that aims to observe, explore, and investigate

discussion

section of a scientific paper in which the author interprets experimental results, describes how variables may be related, and attempts to explain the phenomenon in question

falsifiable

able to be disproven by experimental results

hypothesis

suggested explanation for an observation, which can be tested

hypothesis-based science

form of science that begins with a specific question and potential testable answers

inductive reasoning

form of logical thinking that uses related observations to arrive at a general conclusion

introduction

opening section of a scientific paper, which provides background information about what was known in the field prior to the research reported in the paper

life science

field of science, such as biology, that studies living things

materials and methods

section of a scientific paper that includes a complete description of the substances, methods, and techniques used by the researchers to gather data

natural science

field of science that is related to the physical world and its phenomena and processes

peer-reviewed manuscript

scientific paper that is reviewed by a scientist's colleagues who are experts in the field of study

physical science

field of science, such as geology, astronomy, physics, and chemistry, that studies nonliving matter

plagiarism

using other people's work or ideas without proper citation, creating the false impression that those are the author's original ideas

results

section of a scientific paper in which the author narrates the experimental findings and presents relevant figures, pictures, diagrams, graphs, and tables, without any further interpretation

review article

paper that summarizes and comments on findings that were published as primary literature

science

knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method

scientific method

method of research with defined steps that include observation, formulation of a hypothesis, testing, and confirming or falsifying the hypothesis

serendipity

fortunate accident or a lucky surprise

theory

tested and confirmed explanation for observations or phenomena

variable

part of an experiment that the experimenter can vary or change

Themes and Concepts of Biology

By the end of this section, you will be able to:

- Identify and describe the properties of life
- Describe the levels of organization among living things
- Recognize and interpret a phylogenetic tree
- List examples of different sub disciplines in biology

Biology is the science that studies life, but what exactly is life? This may sound like a silly question with an obvious response, but it is not always easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet the criteria that biologists use to define life. Consequently, virologists are not biologists, strictly speaking. Similarly, some biologists study the early molecular evolution that gave rise to life; since the events that preceded life are not biological events, these scientists are also excluded from biology in the strict sense of the term.

From its earliest beginnings, biology has wrestled with three questions: What are the shared properties that make something “alive”? And once we know something is alive, how do we find meaningful levels of organization in its structure? And, finally, when faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

Properties of Life

All living organisms share several key characteristics or functions: order, sensitivity or response to the environment, reproduction, adaptation, growth and development, regulation, homeostasis, energy processing, and evolution. When viewed together, these nine characteristics serve to define life.

Order



A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (credit: “Ivengo”/Wikimedia Commons)

Organisms are highly organized, coordinated structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex: inside each cell, atoms make up molecules; these in turn make up cell organelles and other cellular inclusions. In multicellular organisms ([link](#)), similar cells form tissues. Tissues, in turn, collaborate to create organs (body structures with a distinct function). Organs work together to form organ systems.

Sensitivity or Response to Stimuli



The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to normal. (credit: Alex Lomas)

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light, climb on fences and walls, or respond to touch ([\[link\]](#)). Even tiny bacteria can move toward or away from chemicals (a process called *chemotaxis*) or light (*phototaxis*). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

Note:

Link to Learning



Watch [this video](#) to see how plants respond to a stimulus—from opening to light, to wrapping a tendril around a branch, to capturing prey.

Reproduction

Single-celled organisms reproduce by first duplicating their DNA, and then dividing it equally as the cell prepares to divide to form two new cells. Multicellular organisms often produce specialized reproductive germline cells that will form new individuals. When reproduction occurs, genes containing DNA are passed along to an organism's offspring. These genes ensure that the offspring will belong to the same species and will have similar characteristics, such as size and shape.

Growth and Development

Organisms grow and develop following specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young ([link](#)) will grow up to exhibit many of the same characteristics as its parents.



Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics. (credit: Rocky Mountain Feline Rescue)

Regulation

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, respond to stimuli, and cope with environmental stresses. Two examples of internal functions regulated in an organism are nutrient transport and blood flow. Organs (groups of tissues working together) perform specific functions, such as carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

Homeostasis



Polar bears (*Ursus maritimus*) and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (credit: “longhorndave”/Flickr)

In order to function properly, cells need to have appropriate conditions such as proper temperature, pH, and appropriate concentration of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through **homeostasis** (literally, “steady state”)—the ability of an organism to maintain constant internal conditions. For example, an organism needs to regulate body temperature through a process known as thermoregulation. Organisms that live in cold climates, such as the polar bear ([\[link\]](#)), have body structures that help them withstand low temperatures and conserve body heat. Structures that aid in this type of insulation include fur, feathers, blubber, and fat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Energy Processing



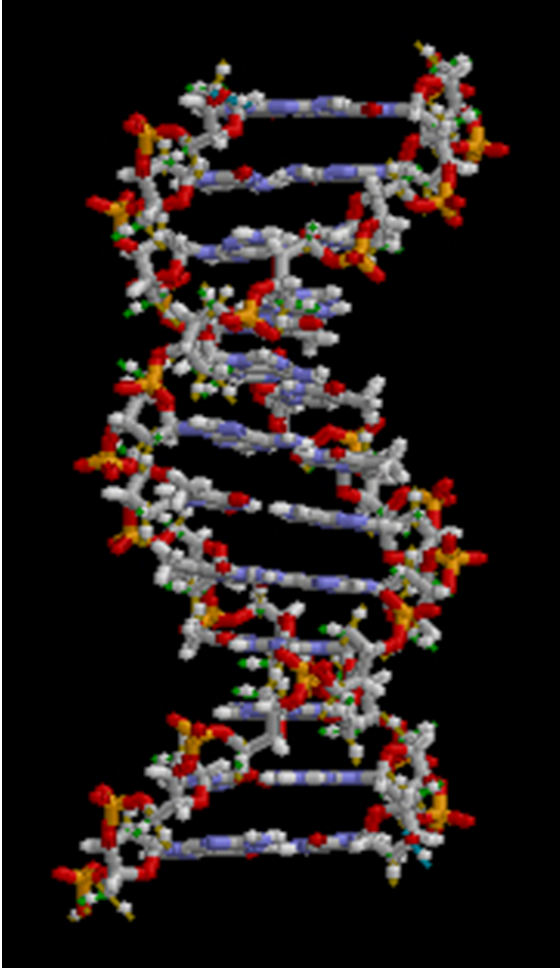
The California condor (*Gymnogyps californianus*) uses chemical energy derived from food to power flight. California condors are an endangered species; this bird has a wing tag that helps biologists identify the individual. (credit: Pacific Southwest Region U.S. Fish and Wildlife Service)

All organisms use a source of energy for their metabolic activities. Some organisms capture energy from the sun and convert it into chemical energy

in food; others use chemical energy in molecules they take in as food ([link](#)).

Levels of Organization of Living Things

Living things are highly organized and structured, following a hierarchy that can be examined on a scale from small to large. The **atom** is the smallest and most fundamental unit of matter. It consists of a nucleus surrounded by electrons. Atoms form molecules. A **molecule** is a chemical structure consisting of at least two atoms held together by one or more chemical bonds. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed by polymerization (a polymer is a large molecule that is made by combining smaller units called monomers, which are simpler than macromolecules). An example of a macromolecule is deoxyribonucleic acid (DNA) ([link](#)), which contains the instructions for the structure and functioning of all living organisms.



All molecules, including this DNA molecule, are composed of atoms. (credit: “brian0918”/Wikimedia Commons)

Note:
Link to Learning



Watch [this video](#) that animates the three-dimensional structure of the DNA molecule shown in [\[link\]](#).

Some cells contain aggregates of macromolecules surrounded by membranes; these are called **organelles**. Organelles are small structures that exist within cells. Examples of organelles include mitochondria and chloroplasts, which carry out indispensable functions: mitochondria produce energy to power the cell, while chloroplasts enable green plants to utilize the energy in sunlight to make sugars. All living things are made of cells; the **cell** itself is the smallest fundamental unit of structure and function in living organisms. (This requirement is why viruses are not considered living: they are not made of cells. To make new viruses, they have to invade and hijack the reproductive mechanism of a living cell; only then can they obtain the materials they need to reproduce.) Some organisms consist of a single cell and others are multicellular. Cells are classified as prokaryotic or eukaryotic. **Prokaryotes** are single-celled or colonial organisms that do not have membrane-bound nuclei; in contrast, the cells of **eukaryotes** do have membrane-bound organelles and a membrane-bound nucleus.

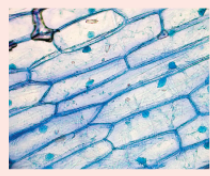
In larger organisms, cells combine to make **tissues**, which are groups of similar cells carrying out similar or related functions. **Organs** are collections of tissues grouped together performing a common function. Organs are present not only in animals but also in plants. An **organ system** is a higher level of organization that consists of functionally related organs. Mammals have many organ systems. For instance, the circulatory system transports blood through the body and to and from the lungs; it includes organs such as the heart and blood vessels. **Organisms** are individual living entities. For example, each tree in a forest is an organism. Single-celled

prokaryotes and single-celled eukaryotes are also considered organisms and are typically referred to as microorganisms.

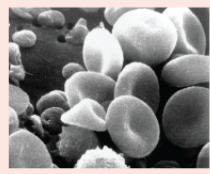
All the individuals of a species living within a specific area are collectively called a **population**. For example, a forest may include many pine trees. All of these pine trees represent the population of pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations. A **community** is the sum of populations inhabiting a particular area. For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community. The forest itself is an ecosystem. An **ecosystem** consists of all the living things in a particular area together with the abiotic, non-living parts of that environment such as nitrogen in the soil or rain water. At the highest level of organization ([\[link\]](#)), the **biosphere** is the collection of all ecosystems, and it represents the zones of life on earth. It includes land, water, and even the atmosphere to a certain extent.

Note:

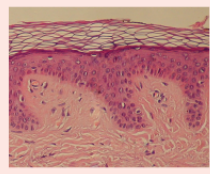
Art Connection



Organelles: The nucleus, dyed blue in these onion cells, is an example of an organelle.



Cells: Human blood cells.



Tissues: Human skin tissue.



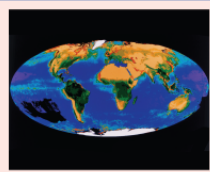
Organs and Organ Systems: Organs, such as the stomach and intestine, make up the human digestive system.



Organisms, Populations, and Communities: In a forest, each pine tree is an organism. Together, all the pine trees make up a population. All the plant and animal species in the forest comprise a community.



Ecosystems: This coastal ecosystem in the southeastern United States includes living organisms and the environment in which they live.



The Biosphere: Encompasses all the ecosystems on Earth.

The biological levels of organization of living things are shown. From a single organelle to the

entire biosphere, living organisms are parts of a highly structured hierarchy. (credit “organelles”: modification of work by Umberto Salvagnin; credit “cells”: modification of work by Bruce Wetzel, Harry Schaefer/ National Cancer Institute; credit “tissues”: modification of work by Kilbad; Fama Clamosa; Mikael Häggström; credit “organs”: modification of work by Mariana Ruiz Villareal; credit “organisms”: modification of work by "Crystal"/Flickr; credit “ecosystems”: modification of work by US Fish and Wildlife Service Headquarters; credit “biosphere”: modification of work by NASA)

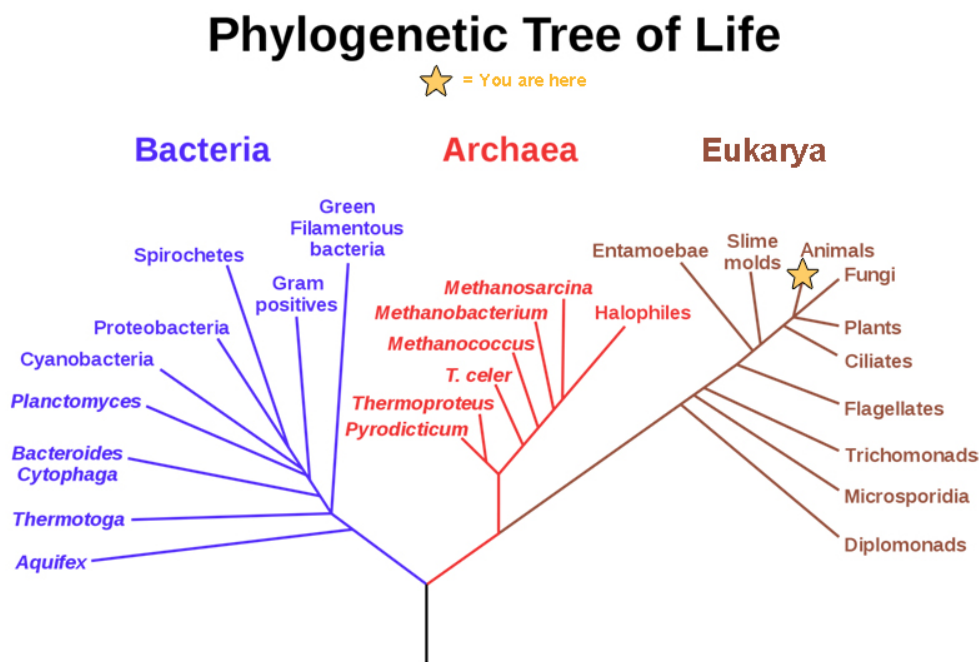
Which of the following statements is false?

- a. Tissues exist within organs which exist within organ systems.
- b. Communities exist within populations which exist within ecosystems.
- c. Organelles exist within cells which exist within tissues.
- d. Communities exist within ecosystems which exist in the biosphere.

The Diversity of Life

The fact that biology, as a science, has such a broad scope has to do with the tremendous diversity of life on earth. The source of this diversity is **evolution**, the process of gradual change during which new species arise from older species. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems.

The evolution of various life forms on Earth can be summarized in a phylogenetic tree ([link](#)). A **phylogenetic tree** is a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both. A phylogenetic tree is composed of nodes and branches. The internal nodes represent ancestors and are points in evolution when, based on scientific evidence, an ancestor is thought to have diverged to form two new species. The length of each branch is proportional to the time elapsed since the split.



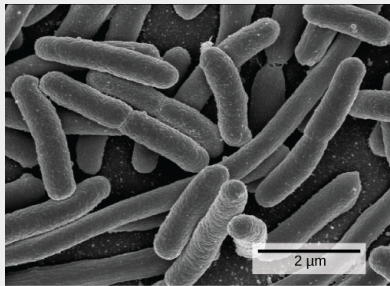
This phylogenetic tree was constructed by microbiologist Carl Woese using data obtained from sequencing ribosomal RNA genes. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are prokaryotes, single-celled organisms lacking intracellular organelles. (credit: Eric Gaba; NASA Astrobiology Institute)

Note:**Evolution Connection****Carl Woese and the Phylogenetic Tree**

In the past, biologists grouped living organisms into five kingdoms: animals, plants, fungi, protists, and bacteria. The organizational scheme was based mainly on physical features, as opposed to physiology, biochemistry, or molecular biology, all of which are used by modern systematics. The pioneering work of American microbiologist Carl Woese in the early 1970s has shown, however, that life on Earth has evolved along three lineages, now called domains—Bacteria, Archaea, and Eukarya. The first two are prokaryotic cells with microbes that lack membrane-enclosed nuclei and organelles. The third domain contains the eukaryotes and includes unicellular microorganisms together with the four original kingdoms (excluding bacteria). Woese defined Archaea as a new domain, and this resulted in a new taxonomic tree ([\[link\]](#)). Many organisms belonging to the Archaea domain live under extreme conditions and are called extremophiles. To construct his tree, Woese used genetic relationships rather than similarities based on morphology (shape). Woese's tree was constructed from comparative sequencing of the genes that are universally distributed, present in every organism, and conserved (meaning that these genes have remained essentially unchanged throughout evolution). Woese's approach was revolutionary because comparisons of physical features are insufficient to differentiate between the prokaryotes that appear fairly similar in spite of their tremendous biochemical diversity and genetic variability ([\[link\]](#)). The comparison of homologous DNA and

RNA sequences provided Woese with a sensitive device that revealed the extensive variability of prokaryotes, and which justified the separation of the prokaryotes into two domains: bacteria and archaea.

These images represent different domains. The (a) bacteria in this micrograph belong to Domain Bacteria, while the (b) extremophiles (not visible) living in this hot vent belong to Domain Archaea. Both the (c) sunflower and (d) lion are part of Domain Eukarya. (credit a: modification of work by Drew March; credit b: modification of work by Steve Jurvetson; credit c: modification of work by Michael Arrighi; credit d: modification of work by Leszek Leszcynski)



(a)



(b)



(c)



(d)

Branches of Biological Study

The scope of biology is broad and therefore contains many branches and subdisciplines. Biologists may pursue one of those subdisciplines and work

in a more focused field. For instance, **molecular biology** and **biochemistry** study biological processes at the molecular and chemical level, including interactions among molecules such as DNA, RNA, and proteins, as well as the way they are regulated. **Microbiology**, the study of microorganisms, is the study of the structure and function of single-celled organisms. It is quite a broad branch itself, and depending on the subject of study, there are also microbial physiologists, ecologists, and geneticists, among others.

Note:**Career Connection****Forensic Scientist**

Forensic science is the application of science to answer questions related to the law. Biologists as well as chemists and biochemists can be forensic scientists. Forensic scientists provide scientific evidence for use in courts, and their job involves examining trace materials associated with crimes. Interest in forensic science has increased in the last few years, possibly because of popular television shows that feature forensic scientists on the job. Also, the development of molecular techniques and the establishment of DNA databases have expanded the types of work that forensic scientists can do. Their job activities are primarily related to crimes against people such as murder, rape, and assault. Their work involves analyzing samples such as hair, blood, and other body fluids and also processing DNA ([\[link\]](#)) found in many different environments and materials. Forensic scientists also analyze other biological evidence left at crime scenes, such as insect larvae or pollen grains. Students who want to pursue careers in forensic science will most likely be required to take chemistry and biology courses as well as some intensive math courses.



This forensic scientist works in a DNA extraction room at the U.S. Army Criminal Investigation Laboratory at Fort Gillem, GA. (credit: United States Army CID Command Public Affairs)

Another field of biological study, **neurobiology**, studies the biology of the nervous system, and although it is considered a branch of biology, it is also recognized as an interdisciplinary field of study known as neuroscience. Because of its interdisciplinary nature, this subdiscipline studies different functions of the nervous system using molecular, cellular, developmental, medical, and computational approaches.



Researchers work on excavating dinosaur fossils at a site in Castellón, Spain. (credit: Mario Modesto)

Paleontology, another branch of biology, uses fossils to study life's history ([link](#)). **Zoology** and **botany** are the study of animals and plants, respectively. Biologists can also specialize as biotechnologists, ecologists, or physiologists, to name just a few areas. This is just a small sample of the many fields that biologists can pursue.

Biology is the culmination of the achievements of the natural sciences from their inception to today. Excitingly, it is the cradle of emerging sciences, such as the biology of brain activity, genetic engineering of custom organisms, and the biology of evolution that uses the laboratory tools of molecular biology to retrace the earliest stages of life on earth. A scan of news headlines—whether reporting on immunizations, a newly discovered species, sports doping, or a genetically-modified food—demonstrates the way biology is active in and important to our everyday world.

Section Summary

Biology is the science of life. All living organisms share several key properties such as order, sensitivity or response to stimuli, reproduction, growth and development, regulation, homeostasis, and energy processing. Living things are highly organized parts of a hierarchy that includes atoms, molecules, organelles, cells, tissues, organs, and organ systems. Organisms, in turn, are grouped as populations, communities, ecosystems, and the biosphere. The great diversity of life today evolved from less-diverse ancestral organisms over billions of years. A diagram called a phylogenetic tree can be used to show evolutionary relationships among organisms.

Biology is very broad and includes many branches and subdisciplines. Examples include molecular biology, microbiology, neurobiology, zoology, and botany, among others.

Art Connections

Exercise:

Problem: [\[link\]](#) Which of the following statements is false?

- a. Tissues exist within organs which exist within organ systems.
- b. Communities exist within populations which exist within ecosystems.
- c. Organelles exist within cells which exist within tissues.
- d. Communities exist within ecosystems which exist in the biosphere.

Solution:

[\[link\]](#) Communities exist within populations which exist within ecosystems.

Review Questions

Exercise:

Problem:

The smallest unit of biological structure that meets the functional requirements of “living” is the _____.

- a. organ
- b. organelle
- c. cell
- d. macromolecule

Solution:

C

Exercise:

Problem: Viruses are not considered living because they _____.

- a. are not made of cells
- b. lack cell nuclei
- c. do not contain DNA or RNA
- d. cannot reproduce

Solution:

A

Exercise:**Problem:**

The presence of a membrane-enclosed nucleus is a characteristic of _____.

- a. prokaryotic cells
- b. eukaryotic cells
- c. living organisms

d. bacteria

Solution:

B

Exercise:

Problem:

A group of individuals of the same species living in the same area is called a(n) _____.

- a. family
- b. community
- c. population
- d. ecosystem

Solution:

C

Exercise:

Problem:

Which of the following sequences represents the hierarchy of biological organization from the most inclusive to the least complex level?

- a. organelle, tissue, biosphere, ecosystem, population
- b. organ, organism, tissue, organelle, molecule
- c. organism, community, biosphere, molecule, tissue, organ
- d. biosphere, ecosystem, community, population, organism

Solution:

D

Exercise:**Problem:**

Where in a phylogenetic tree would you expect to find the organism that had evolved most recently?

- a. at the base
- b. within the branches
- c. at the nodes
- d. at the branch tips

Solution:

D

Free Response**Exercise:****Problem:**

Select two items that biologists agree are necessary in order to consider an organism “alive.” For each, give an example of a non-living object that otherwise fits the definition of “alive,”

Solution:

Answers will vary. Layers of sedimentary rock have order but are not alive. Technology is capable of regulation but is not, of itself, alive.

Exercise:

Problem:

Consider the levels of organization of the biological world, and place each of these items in order from smallest level of organization to most encompassing: skin cell, elephant, water molecule, planet Earth, tropical rainforest, hydrogen atom, wolf pack, liver.

Solution:

Smallest level of organization to largest: hydrogen atom, water molecule, skin cell, liver, elephant, wolf pack, tropical rainforest, planet Earth

Exercise:**Problem:**

You go for a long walk on a hot day. Give an example of a way in which homeostasis keeps your body healthy.

Solution:

During your walk, you may begin to perspire, which cools your body and helps your body to maintain a constant internal temperature. You might also become thirsty and pause long enough for a cool drink, which will help to restore the water lost during perspiration.

Exercise:**Problem:**

Using examples, explain how biology can be studied from a microscopic approach to a global approach.

Solution:

Researchers can approach biology from the smallest to the largest, and everything in between. For instance, an ecologist may study a population of individuals, the population's community, the community's ecosystem, and the ecosystem's part in the biosphere.

When studying an individual organism, a biologist could examine the cell and its organelles, the tissues that the cells make up, the organs and their respective organ systems, and the sum total—the organism itself.

Glossary

atom

smallest and most fundamental unit of matter

biochemistry

study of the chemistry of biological organisms

biosphere

collection of all the ecosystems on Earth

botany

study of plants

cell

smallest fundamental unit of structure and function in living things

community

set of populations inhabiting a particular area

ecosystem

all the living things in a particular area together with the abiotic, nonliving parts of that environment

eukaryote

organism with cells that have nuclei and membrane-bound organelles

evolution

process of gradual change during which new species arise from older species and some species become extinct

homeostasis

ability of an organism to maintain constant internal conditions

macromolecule

large molecule, typically formed by the joining of smaller molecules

microbiology

study of the structure and function of microorganisms

molecule

chemical structure consisting of at least two atoms held together by one or more chemical bonds

molecular biology

study of biological processes and their regulation at the molecular level, including interactions among molecules such as DNA, RNA, and proteins

neurobiology

study of the biology of the nervous system

organ

collection of related tissues grouped together performing a common function

organ system

level of organization that consists of functionally related interacting organs

organelle

small structures that exist within cells and carry out cellular functions

organism

individual living entity

paleontology

study of life's history by means of fossils

phylogenetic tree

diagram showing the evolutionary relationships among various biological species based on similarities and differences in genetic or physical traits or both; in essence, a hypothesis concerning evolutionary connections

population

all of the individuals of a species living within a specific area

prokaryote

single-celled organism that lacks organelles and does not have nuclei surrounded by a nuclear membrane

tissue

group of similar cells carrying out related functions

zoology

study of animals

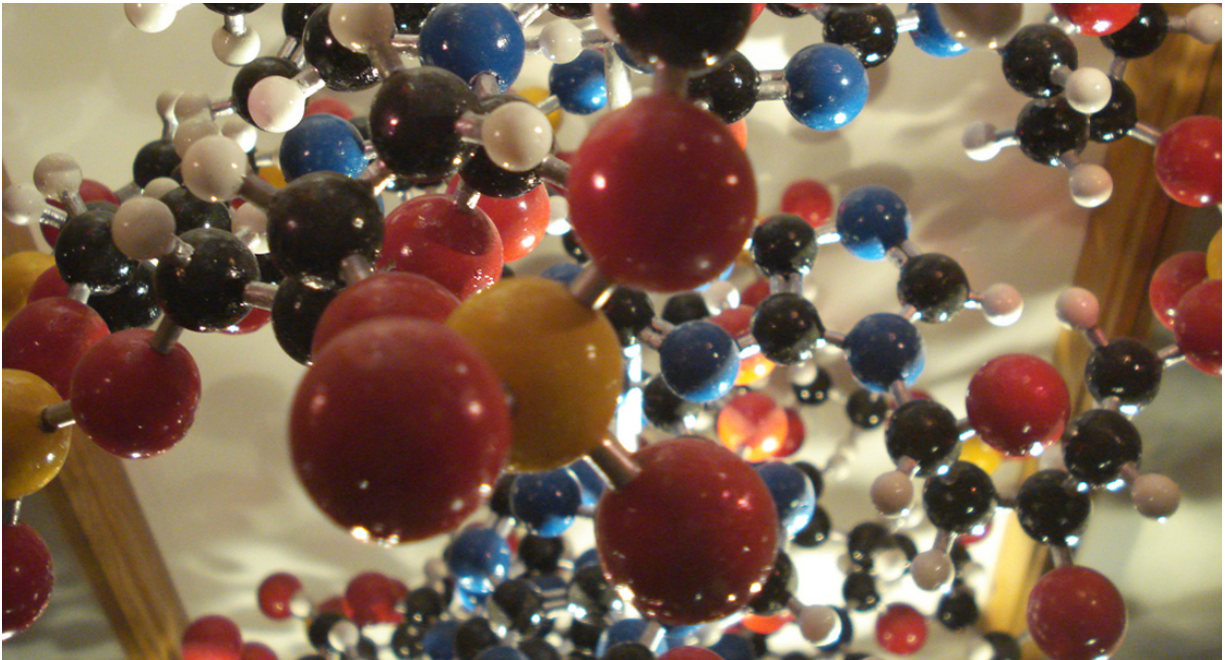
Introduction

class="introduction"

Atoms are
the building
blocks of
molecules
found in the
universe—
air, soil,
water, rocks
. . . and also
the cells of
all living
organisms.

In this
model of an
organic
molecule,
the atoms of
carbon
(black),
hydrogen
(white),
nitrogen
(blue),
oxygen
(red), and
sulfur
(yellow) are
shown in
proportional
atomic size.
The silver
rods
indicate
chemical

bonds.
(credit:
modification
n of work
by Christian
Guthier)



Elements in various combinations comprise all matter, including living things. Some of the most abundant elements in living organisms include carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus. These form the nucleic acids, proteins, carbohydrates, and lipids that are the fundamental components of living matter. Biologists must understand these important building blocks and the unique structures of the atoms that make up molecules, allowing for the formation of cells, tissues, organ systems, and entire organisms.

All biological processes follow the laws of physics and chemistry, so in order to understand how biological systems work, it is important to understand the underlying physics and chemistry. For example, the flow of blood within the circulatory system follows the laws of physics that regulate the modes of fluid flow. The breakdown of the large, complex molecules of

food into smaller molecules—and the conversion of these to release energy to be stored in adenosine triphosphate (ATP)—is a series of chemical reactions that follow chemical laws. The properties of water and the formation of hydrogen bonds are key to understanding living processes. Recognizing the properties of acids and bases is important, for example, to our understanding of the digestive process. Therefore, the fundamentals of physics and chemistry are important for gaining insight into biological processes.

Atoms, Isotopes, Ions, and Molecules: The Building Blocks

By the end of this section, you will be able to:

- Define matter and elements
- Describe the interrelationship between protons, neutrons, and electrons
- Compare the ways in which electrons can be donated or shared between atoms
- Explain the ways in which naturally occurring elements combine to create molecules, cells, tissues, organ systems, and organisms

At its most fundamental level, life is made up of matter. **Matter** is any substance that occupies space and has mass. **Elements** are unique forms of matter with specific chemical and physical properties that cannot be broken down into smaller substances by ordinary chemical reactions. There are 118 elements, but only 92 occur naturally. The remaining elements are synthesized in laboratories and are unstable.

Each element is designated by its chemical symbol, which is a single capital letter or, when the first letter is already “taken” by another element, a combination of two letters. Some elements follow the English term for the element, such as C for carbon and Ca for calcium. Other elements’ chemical symbols derive from their Latin names; for example, the symbol for sodium is Na, referring to *natrium*, the Latin word for sodium.

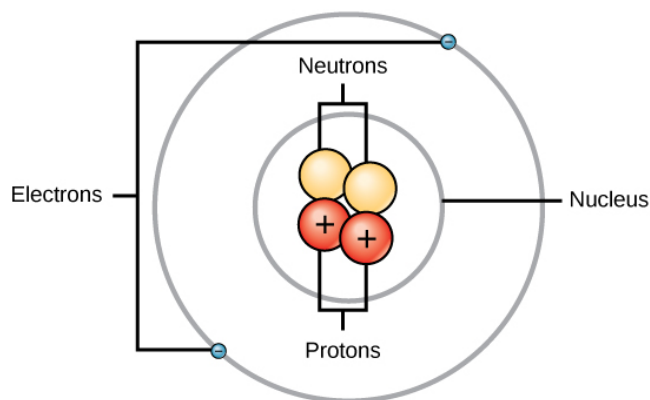
The four elements common to all living organisms are oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). In the non-living world, elements are found in different proportions, and some elements common to living organisms are relatively rare on the earth as a whole, as shown in [\[link\]](#). For example, the atmosphere is rich in nitrogen and oxygen but contains little carbon and hydrogen, while the earth’s crust, although it contains oxygen and a small amount of hydrogen, has little nitrogen and carbon. In spite of their differences in abundance, all elements and the chemical reactions between them obey the same chemical and physical laws regardless of whether they are a part of the living or non-living world.

Approximate Percentage of Elements in Living Organisms (Humans) Compared to the Non-living World			
Element	Life (Humans)	Atmosphere	Earth's Crust
Oxygen (O)	65%	21%	46%
Carbon (C)	18%	trace	trace
Hydrogen (H)	10%	trace	0.1%
Nitrogen (N)	3%	78%	trace

The Structure of the Atom

To understand how elements come together, we must first discuss the smallest component or building block of an element, the atom. An **atom** is the smallest unit of matter that retains all of the chemical properties of an element. For example, one gold atom has all of the properties of gold in that it is a solid metal at room temperature. A gold coin is simply a very large number of gold atoms molded into the shape of a coin and containing small amounts of other elements known as impurities. Gold atoms cannot be broken down into anything smaller while still retaining the properties of gold.

An atom is composed of two regions: the **nucleus**, which is in the center of the atom and contains protons and neutrons, and the outermost region of the atom which holds its electrons in orbit around the nucleus, as illustrated in [\[link\]](#). Atoms contain protons, electrons, and neutrons, among other subatomic particles. The only exception is hydrogen (H), which is made of one proton and one electron with no neutrons.



Elements, such as helium, depicted here, are made up of atoms. Atoms are made up of protons and neutrons located within the nucleus, with electrons in orbitals surrounding the nucleus.

Protons and neutrons have approximately the same mass, about 1.67×10^{-24} grams. Scientists arbitrarily define this amount of mass as one atomic mass unit (amu) or one Dalton, as shown in [\[link\]](#). Although similar in mass, protons and neutrons differ in their electric charge. A **proton** is positively charged whereas a **neutron** is uncharged. Therefore, the number of neutrons in an atom contributes significantly to its mass, but not to its charge. **Electrons** are much smaller in mass than protons, weighing only 9.11×10^{-28} grams, or about 1/1800 of an atomic mass unit. Hence, they do not contribute much to an element's overall atomic mass. Therefore, when considering atomic mass, it is customary to ignore the mass of any electrons and calculate the atom's mass based on the number of protons and neutrons alone. Although not significant contributors to mass, electrons do contribute greatly to the atom's charge, as each electron has a negative charge equal to the positive charge of a proton. In uncharged, neutral atoms, the number of electrons orbiting the nucleus is equal to the number of protons inside the nucleus. In these atoms, the positive and negative charges cancel each other out, leading to an atom with no net charge.

Accounting for the sizes of protons, neutrons, and electrons, most of the volume of an atom—greater than 99 percent—is, in fact, empty space. With all this empty space, one might ask why so-called solid objects do not just pass through one another. The reason they do not is that the electrons that surround all atoms are negatively charged and negative charges repel each other.

Protons, Neutrons, and Electrons			
	Charge	Mass (amu)	Location
Proton	+1	1	nucleus
Neutron	0	1	nucleus
Electron	−1	0	orbitals

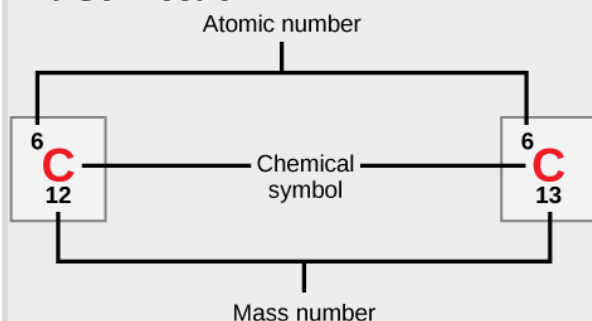
Atomic Number and Mass

Atoms of each element contain a characteristic number of protons and electrons. The number of protons determines an element's **atomic number** and is used to distinguish one element from another. The number of neutrons is variable, resulting in isotopes, which are different forms of the same atom that vary only in the number of neutrons they possess. Together, the number of protons and the number of neutrons determine an element's **mass number**, as illustrated in [\[link\]](#). Note that the small contribution of mass from electrons is disregarded in calculating the mass number. This approximation of mass can be used to easily calculate how many neutrons an element has by simply subtracting the number of protons from the mass number. Since an element's isotopes will have slightly different mass numbers, scientists also determine the **atomic mass**, which is the calculated

mean of the mass number for its naturally occurring isotopes. Often, the resulting number contains a fraction. For example, the atomic mass of chlorine (Cl) is 35.45 because chlorine is composed of several isotopes, some (the majority) with atomic mass 35 (17 protons and 18 neutrons) and some with atomic mass 37 (17 protons and 20 neutrons).

Note:

Art Connection



Carbon has an atomic number of six, and two stable isotopes with mass numbers of twelve and thirteen, respectively. Its relative atomic mass is 12.011

How many neutrons do carbon-12 and carbon-13 have, respectively?

Isotopes

Isotopes are different forms of an element that have the same number of protons but a different number of neutrons. Some elements—such as carbon, potassium, and uranium—have naturally occurring isotopes. Carbon-12 contains six protons, six neutrons, and six electrons; therefore, it has a mass number of 12 (six protons and six neutrons). Carbon-14 contains six protons, eight neutrons, and six electrons; its atomic mass is 14 (six

protons and eight neutrons). These two alternate forms of carbon are isotopes. Some isotopes may emit neutrons, protons, and electrons, and attain a more stable atomic configuration (lower level of potential energy); these are radioactive isotopes, or **radioisotopes**. Radioactive decay (carbon-14 decaying to eventually become nitrogen-14) describes the energy loss that occurs when an unstable atom's nucleus releases radiation.

Note:**Evolution Connection****Carbon Dating**

Carbon is normally present in the atmosphere in the form of gaseous compounds like carbon dioxide and methane. Carbon-14 (^{14}C) is a naturally occurring radioisotope that is created in the atmosphere from atmospheric ^{14}N (nitrogen) by the addition of a neutron and the loss of a proton because of cosmic rays. This is a continuous process, so more ^{14}C is always being created. As a living organism incorporates ^{14}C initially as carbon dioxide fixed in the process of photosynthesis, the relative amount of ^{14}C in its body is equal to the concentration of ^{14}C in the atmosphere. When an organism dies, it is no longer ingesting ^{14}C , so the ratio between ^{14}C and ^{12}C will decline as ^{14}C decays gradually to ^{14}N by a process called beta decay—the emission of electrons or positrons. This decay gives off energy in a slow process.

After approximately 5,730 years, half of the starting concentration of ^{14}C will have been converted back to ^{14}N . The time it takes for half of the original concentration of an isotope to decay back to its more stable form is called its half-life. Because the half-life of ^{14}C is long, it is used to date formerly living objects such as old bones or wood. Comparing the ratio of the ^{14}C concentration found in an object to the amount of ^{14}C detected in the atmosphere, the amount of the isotope that has not yet decayed can be determined. On the basis of this amount, the age of the material, such as the pygmy mammoth shown in [\[link\]](#), can be calculated with accuracy if it is not much older than about 50,000 years. Other elements have isotopes with different half lives. For example, ^{40}K (potassium-40) has a half-life of 1.25 billion years, and ^{235}U (Uranium 235) has a half-life of about 700 million years. Through the use of radiometric dating, scientists can study

the age of fossils or other remains of extinct organisms to understand how organisms have evolved from earlier species.



The age of carbon-containing remains less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating. (credit: Bill Faulkner, NPS)

Note:

Link to Learning



To learn more about atoms, isotopes, and how to tell one isotope from another, run the simulation.

https://openstax.org/l/atoms_isotopes

The Periodic Table

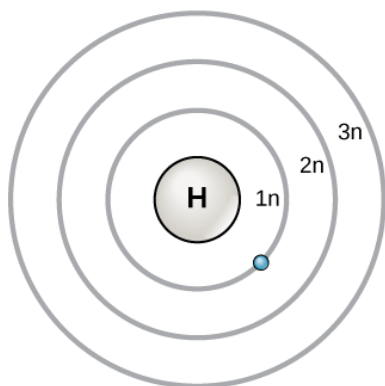
The different elements are organized and displayed in the **periodic table**. Devised by Russian chemist Dmitri Mendeleev (1834–1907) in 1869, the table groups elements that, although unique, share certain chemical properties with other elements. The properties of elements are responsible for their physical state at room temperature: they may be gases, solids, or liquids. Elements also have specific **chemical reactivity**, the ability to combine and to chemically bond with each other.

In the periodic table, shown in [\[link\]](#), the elements are organized and displayed according to their atomic number and are arranged in a series of rows and columns based on shared chemical and physical properties. In addition to providing the atomic number for each element, the periodic table also displays the element's atomic mass. Looking at carbon, for example, its symbol (C) and name appear, as well as its atomic number of six (in the upper left-hand corner) and its atomic mass of 12.11.

Electron Shells and the Bohr Model

It should be stressed that there is a connection between the number of protons in an element, the atomic number that distinguishes one element from another, and the number of electrons it has. In all electrically neutral atoms, the number of electrons is the same as the number of protons. Thus, each element, at least when electrically neutral, has a characteristic number of electrons equal to its atomic number.

An early model of the atom was developed in 1913 by Danish scientist Niels Bohr (1885–1962). The Bohr model shows the atom as a central nucleus containing protons and neutrons, with the electrons in circular **orbitals** at specific distances from the nucleus, as illustrated in [\[link\]](#). These orbits form electron shells or energy levels, which are a way of visualizing the number of electrons in the outermost shells. These energy levels are designated by a number and the symbol “n.” For example, 1n represents the first energy level located closest to the nucleus.



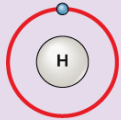
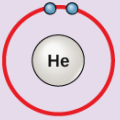
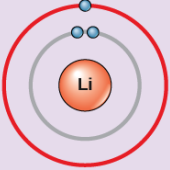
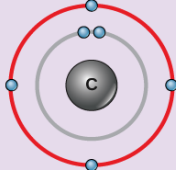
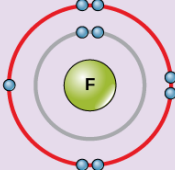
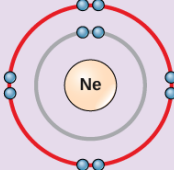
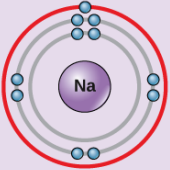
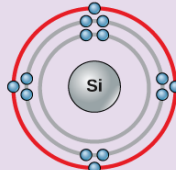
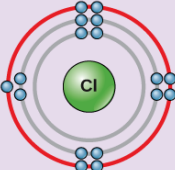
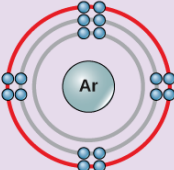
The Bohr model was developed by Niels Bohrs in 1913. In this model, electrons exist within principal shells. An electron normally exists

in the lowest energy shell available, which is the one closest to the nucleus. Energy from a photon of light can bump it up to a higher energy shell, but this situation is unstable, and the electron quickly decays back to the ground state. In the process, a photon of light is released.

Electrons fill orbitals in a consistent order: they first fill the orbitals closest to the nucleus, then they continue to fill orbitals of increasing energy further from the nucleus. If there are multiple orbitals of equal energy, they will be filled with one electron in each energy level before a second electron is added. The electrons of the outermost energy level determine the energetic stability of the atom and its tendency to form chemical bonds with other atoms to form molecules.

Under standard conditions, atoms fill the inner shells first, often resulting in a variable number of electrons in the outermost shell. The innermost shell has a maximum of two electrons but the next two electron shells can each have a maximum of eight electrons. This is known as the **octet rule**, which states, with the exception of the innermost shell, that atoms are more stable energetically when they have eight electrons in their **valence shell**, the outermost electron shell. Examples of some neutral atoms and their electron configurations are shown in [\[link\]](#). Notice that in this [\[link\]](#), helium has a complete outer electron shell, with two electrons filling its first and only shell. Similarly, neon has a complete outer $2n$ shell containing eight electrons. In contrast, chlorine and sodium have seven and one in their outer shells, respectively, but theoretically they would be more energetically stable if they followed the octet rule and had eight.

Note:**Art Connection**

	Group 1	Group 14	Group 17	Group 18
Period 1 (1n is filling)				
Period 2 (2n is filling)				
Period 3 (3n is filling)				

Bohr diagrams indicate how many electrons fill each principal shell. Group 18 elements (helium, neon, and argon are shown) have a full outer, or valence, shell. A full valence shell is the most stable electron configuration. Elements in other groups have partially filled valence shells and gain or lose electrons to achieve a stable electron configuration.

An atom may give, take, or share electrons with another atom to achieve a full valence shell, the most stable electron configuration. Looking at this figure, how many electrons do elements in group 1 need to lose in order to achieve a stable electron configuration? How many electrons do elements in groups 14 and 17 need to gain to achieve a stable configuration?

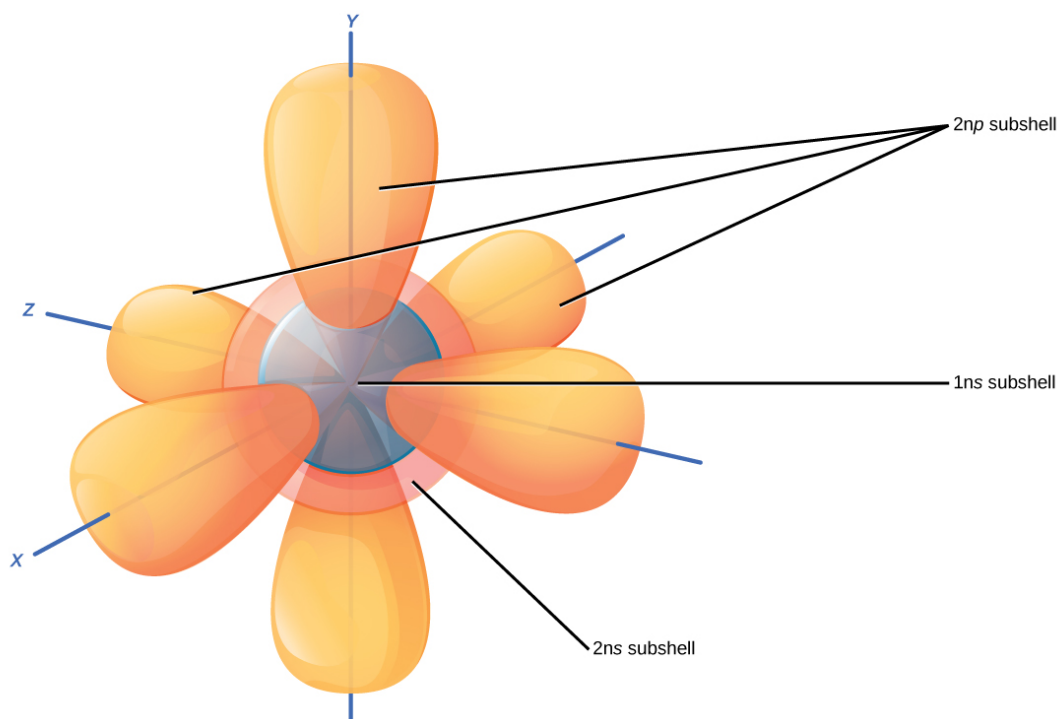
Understanding that the organization of the periodic table is based on the total number of protons (and electrons) helps us know how electrons are distributed among the outer shell. The periodic table is arranged in columns

and rows based on the number of electrons and where these electrons are located. Take a closer look at the some of the elements in the table's far right column in [\[link\]](#). The group 18 atoms helium (He), neon (Ne), and argon (Ar) all have filled outer electron shells, making it unnecessary for them to share electrons with other atoms to attain stability; they are highly stable as single atoms. Their non-reactivity has resulted in their being named the **inert gases** (or **noble gases**). Compare this to the group 1 elements in the left-hand column. These elements, including hydrogen (H), lithium (Li), and sodium (Na), all have one electron in their outermost shells. That means that they can achieve a stable configuration and a filled outer shell by donating or sharing one electron with another atom or a molecule such as water. Hydrogen will donate or share its electron to achieve this configuration, while lithium and sodium will donate their electron to become stable. As a result of losing a negatively charged electron, they become positively charged **ions**. Group 17 elements, including fluorine and chlorine, have seven electrons in their outmost shells, so they tend to fill this shell with an electron from other atoms or molecules, making them negatively charged ions. Group 14 elements, of which carbon is the most important to living systems, have four electrons in their outer shell allowing them to make several covalent bonds (discussed below) with other atoms. Thus, the columns of the periodic table represent the potential shared state of these elements' outer electron shells that is responsible for their similar chemical characteristics.

Electron Orbitals

Although useful to explain the reactivity and chemical bonding of certain elements, the Bohr model of the atom does not accurately reflect how electrons are spatially distributed surrounding the nucleus. They do not circle the nucleus like the earth orbits the sun, but are found in **electron orbitals**. These relatively complex shapes result from the fact that electrons behave not just like particles, but also like waves. Mathematical equations from quantum mechanics known as wave functions can predict within a certain level of probability where an electron might be at any given time. The area where an electron is most likely to be found is called its orbital.

Recall that the Bohr model depicts an atom's electron shell configuration. Within each electron shell are subshells, and each subshell has a specified number of orbitals containing electrons. While it is impossible to calculate exactly where an electron is located, scientists know that it is most probably located within its orbital path. Subshells are designated by the letter *s*, *p*, *d*, and *f*. The *s* subshell is spherical in shape and has one orbital. Principal shell $1n$ has only a single *s* orbital, which can hold two electrons. Principal shell $2n$ has one *s* and one *p* subshell, and can hold a total of eight electrons. The *p* subshell has three dumbbell-shaped orbitals, as illustrated in [\[link\]](#). Subshells *d* and *f* have more complex shapes and contain five and seven orbitals, respectively. These are not shown in the illustration. Principal shell $3n$ has *s*, *p*, and *d* subshells and can hold 18 electrons. Principal shell $4n$ has *s*, *p*, *d* and *f* orbitals and can hold 32 electrons. Moving away from the nucleus, the number of electrons and orbitals found in the energy levels increases. Progressing from one atom to the next in the periodic table, the electron structure can be worked out by fitting an extra electron into the next available orbital.



The *s* subshells are shaped like spheres. Both the $1n$ and $2n$

principal shells have an *s* orbital, but the size of the sphere is larger in the $2n$ orbital. Each sphere is a single orbital. *p* subshells are made up of three dumbbell-shaped orbitals. Principal shell $2n$ has a *p* subshell, but shell 1 does not.

The closest orbital to the nucleus, called the $1s$ orbital, can hold up to two electrons. This orbital is equivalent to the innermost electron shell of the Bohr model of the atom. It is called the $1s$ orbital because it is spherical around the nucleus. The $1s$ orbital is the closest orbital to the nucleus, and it is always filled first, before any other orbital can be filled. Hydrogen has one electron; therefore, it has only one spot within the $1s$ orbital occupied. This is designated as $1s^1$, where the superscripted 1 refers to the one electron within the $1s$ orbital. Helium has two electrons; therefore, it can completely fill the $1s$ orbital with its two electrons. This is designated as $1s^2$, referring to the two electrons of helium in the $1s$ orbital. On the periodic table [\[link\]](#), hydrogen and helium are the only two elements in the first row (period); this is because they only have electrons in their first shell, the $1s$ orbital. Hydrogen and helium are the only two elements that have the $1s$ and no other electron orbitals in the electrically neutral state.

The second electron shell may contain eight electrons. This shell contains another spherical *s* orbital and three “dumbbell” shaped *p* orbitals, each of which can hold two electrons, as shown in [\[link\]](#). After the $1s$ orbital is filled, the second electron shell is filled, first filling its $2s$ orbital and then its three *p* orbitals. When filling the *p* orbitals, each takes a single electron; once each *p* orbital has an electron, a second may be added. Lithium (Li) contains three electrons that occupy the first and second shells. Two electrons fill the $1s$ orbital, and the third electron then fills the $2s$ orbital. Its **electron configuration** is $1s^2 2s^1$. Neon (Ne), on the other hand, has a total of ten electrons: two are in its innermost $1s$ orbital and eight fill its second shell (two each in the $2s$ and three *p* orbitals); thus, it is an inert gas and energetically stable as a single atom that will rarely form a chemical bond with other atoms. Larger elements have additional orbitals, making up the third electron shell. While the concepts of electron shells and orbitals are closely related, orbitals provide a more accurate depiction of the electron

configuration of an atom because the orbital model specifies the different shapes and special orientations of all the places that electrons may occupy.

Note:

Link to Learning

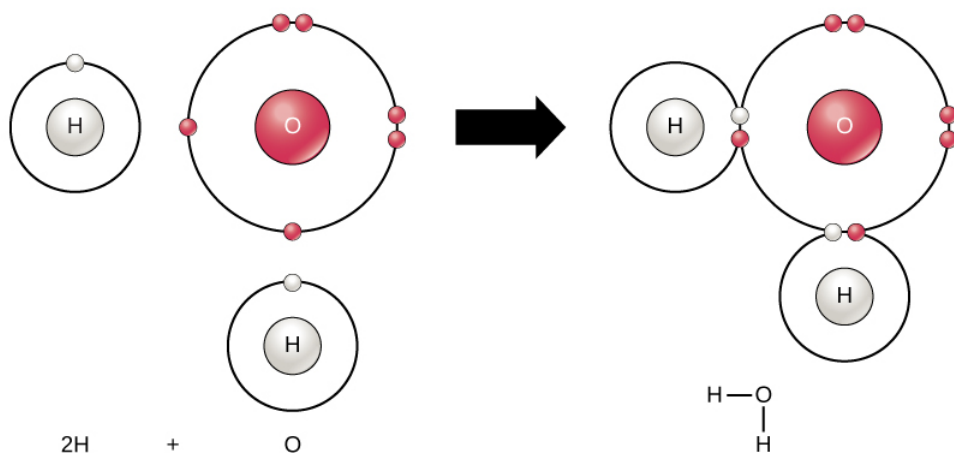


Watch this visual animation to see the spatial arrangement of the p and s orbitals.

<https://www.openstaxcollege.org/l/orbitals>

Chemical Reactions and Molecules

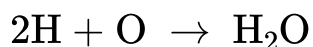
All elements are most stable when their outermost shell is filled with electrons according to the octet rule. This is because it is energetically favorable for atoms to be in that configuration and it makes them stable. However, since not all elements have enough electrons to fill their outermost shells, atoms form **chemical bonds** with other atoms thereby obtaining the electrons they need to attain a stable electron configuration. When two or more atoms chemically bond with each other, the resultant chemical structure is a molecule. The familiar water molecule, H_2O , consists of two hydrogen atoms and one oxygen atom; these bond together to form water, as illustrated in [\[link\]](#). Atoms can form molecules by donating, accepting, or sharing electrons to fill their outer shells.



Two or more atoms may bond with each other to form a molecule. When two hydrogens and an oxygen share electrons via covalent bonds, a water molecule is formed.

Chemical reactions occur when two or more atoms bond together to form molecules or when bonded atoms are broken apart. The substances used in the beginning of a chemical reaction are called the **reactants** (usually found on the left side of a chemical equation), and the substances found at the end of the reaction are known as the **products** (usually found on the right side of a chemical equation). An arrow is typically drawn between the reactants and products to indicate the direction of the chemical reaction; this direction is not always a “one-way street.” For the creation of the water molecule shown above, the chemical equation would be:

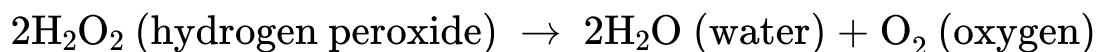
Equation:



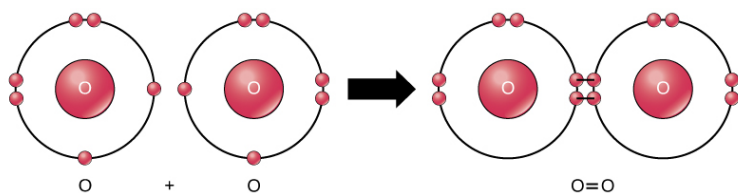
An example of a simple chemical reaction is the breaking down of hydrogen peroxide molecules, each of which consists of two hydrogen atoms bonded to two oxygen atoms (H_2O_2). The reactant hydrogen peroxide is broken down into water, containing one oxygen atom bound to two hydrogen atoms (H_2O), and oxygen, which consists of two bonded oxygen atoms (O_2). In the equation below, the reaction includes two

hydrogen peroxide molecules and two water molecules. This is an example of a **balanced chemical equation**, wherein the number of atoms of each element is the same on each side of the equation. According to the law of conservation of matter, the number of atoms before and after a chemical reaction should be equal, such that no atoms are, under normal circumstances, created or destroyed.

Equation:



Even though all of the reactants and products of this reaction are molecules (each atom remains bonded to at least one other atom), in this reaction only hydrogen peroxide and water are representatives of **compounds**: they contain atoms of more than one type of element. Molecular oxygen, on the other hand, as shown in [\[link\]](#), consists of two doubly bonded oxygen atoms and is not classified as a compound but as a homonuclear molecule.



The oxygen atoms in an O_2 molecule are joined by a double bond.

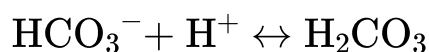
Some chemical reactions, such as the one shown above, can proceed in one direction until the reactants are all used up. The equations that describe these reactions contain a unidirectional arrow and are **irreversible**.

Reversible reactions are those that can go in either direction. In reversible reactions, reactants are turned into products, but when the concentration of product goes beyond a certain threshold (characteristic of the particular reaction), some of these products will be converted back into reactants; at this point, the designations of products and reactants are reversed. This back

and forth continues until a certain relative balance between reactants and products occurs—a state called **equilibrium**. These situations of reversible reactions are often denoted by a chemical equation with a double headed arrow pointing towards both the reactants and products.

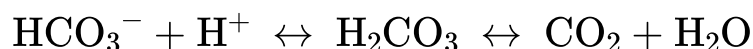
For example, in human blood, excess hydrogen ions (H^+) bind to bicarbonate ions (HCO_3^-) forming an equilibrium state with carbonic acid (H_2CO_3). If carbonic acid were added to this system, some of it would be converted to bicarbonate and hydrogen ions.

Equation:



In biological reactions, however, equilibrium is rarely obtained because the concentrations of the reactants or products or both are constantly changing, often with a product of one reaction being a reactant for another. To return to the example of excess hydrogen ions in the blood, the formation of carbonic acid will be the major direction of the reaction. However, the carbonic acid can also leave the body as carbon dioxide gas (via exhalation) instead of being converted back to bicarbonate ion, thus driving the reaction to the right by the chemical law known as **law of mass action**. These reactions are important for maintaining the homeostasis of our blood.

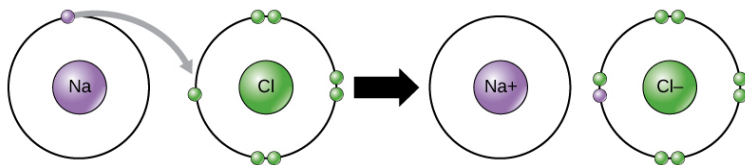
Equation:



Ions and Ionic Bonds

Some atoms are more stable when they gain or lose an electron (or possibly two) and form ions. This fills their outermost electron shell and makes them energetically more stable. Because the number of electrons does not equal the number of protons, each ion has a net charge. **Cations** are positive ions that are formed by losing electrons. Negative ions are formed by gaining electrons and are called anions. **Anions** are designated by their elemental name being altered to end in “-ide”: the anion of chlorine is called chloride, and the anion of sulfur is called sulfide, for example.

This movement of electrons from one element to another is referred to as **electron transfer**. As [\[link\]](#) illustrates, sodium (Na) only has one electron in its outer electron shell. It takes less energy for sodium to donate that one electron than it does to accept seven more electrons to fill the outer shell. If sodium loses an electron, it now has 11 protons, 11 neutrons, and only 10 electrons, leaving it with an overall charge of +1. It is now referred to as a sodium ion. Chlorine (Cl) in its lowest energy state (called the ground state) has seven electrons in its outer shell. Again, it is more energy-efficient for chlorine to gain one electron than to lose seven. Therefore, it tends to gain an electron to create an ion with 17 protons, 17 neutrons, and 18 electrons, giving it a net negative (−1) charge. It is now referred to as a chloride ion. In this example, sodium will donate its one electron to empty its shell, and chlorine will accept that electron to fill its shell. Both ions now satisfy the octet rule and have complete outermost shells. Because the number of electrons is no longer equal to the number of protons, each is now an ion and has a +1 (sodium cation) or −1 (chloride anion) charge. Note that these transactions can normally only take place simultaneously: in order for a sodium atom to lose an electron, it must be in the presence of a suitable recipient like a chlorine atom.



In the formation of an ionic compound, metals lose electrons and nonmetals gain electrons to achieve an octet.

Ionic bonds are formed between ions with opposite charges. For instance, positively charged sodium ions and negatively charged chloride ions bond together to make crystals of sodium chloride, or table salt, creating a crystalline molecule with zero net charge.

Certain salts are referred to in physiology as **electrolytes** (including sodium, potassium, and calcium), ions necessary for nerve impulse conduction, muscle contractions and water balance. Many sports drinks and dietary supplements provide these ions to replace those lost from the body via sweating during exercise.

Covalent Bonds and Other Bonds and Interactions

Another way the octet rule can be satisfied is by the sharing of electrons between atoms to form **covalent bonds**. These bonds are stronger and much more common than ionic bonds in the molecules of living organisms. Covalent bonds are commonly found in carbon-based organic molecules, such as our DNA and proteins. Covalent bonds are also found in inorganic molecules like H_2O , CO_2 , and O_2 . One, two, or three pairs of electrons may be shared, making single, double, and triple bonds, respectively. The more covalent bonds between two atoms, the stronger their connection. Thus, triple bonds are the strongest.

The strength of different levels of covalent bonding is one of the main reasons living organisms have a difficult time in acquiring nitrogen for use in constructing their molecules, even though molecular nitrogen, N_2 , is the most abundant gas in the atmosphere. Molecular nitrogen consists of two nitrogen atoms triple bonded to each other and, as with all molecules, the sharing of these three pairs of electrons between the two nitrogen atoms allows for the filling of their outer electron shells, making the molecule more stable than the individual nitrogen atoms. This strong triple bond makes it difficult for living systems to break apart this nitrogen in order to use it as constituents of proteins and DNA.

The formation of water molecules provides an example of covalent bonding. The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds, as shown in [\[link\]](#). The electron from the hydrogen splits its time between the incomplete outer shell of the hydrogen atoms and the incomplete outer shell of the oxygen atoms. To completely fill the outer shell of oxygen, which has six electrons in its outer shell but which would be more stable with eight, two electrons (one from each hydrogen atom) are needed: hence the well-known formula

H₂O. The electrons are shared between the two elements to fill the outer shell of each, making both elements more stable.

Note:

Link to Learning



View this short video to see an animation of ionic and covalent bonding.
https://www.openstaxcollege.org/l/ionic_covalent

Polar Covalent Bonds

There are two types of covalent bonds: polar and nonpolar. In a **polar covalent bond**, shown in [\[link\]](#), the electrons are unequally shared by the atoms and are attracted more to one nucleus than the other. Because of the unequal distribution of electrons between the atoms of different elements, a slightly positive (δ^+) or slightly negative (δ^-) charge develops. This partial charge is an important property of water and accounts for many of its characteristics.

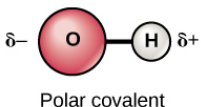
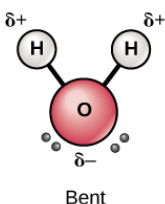
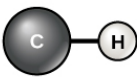
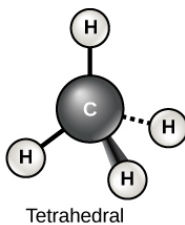
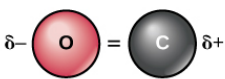

Water is a polar molecule, with the hydrogen atoms acquiring a partial positive charge and the oxygen a partial negative charge. This occurs because the nucleus of the oxygen atom is more attractive to the electrons of the hydrogen atoms than the hydrogen nucleus is to the oxygen's electrons. Thus oxygen has a higher **electronegativity** than hydrogen and the shared electrons spend more time in the vicinity of the oxygen nucleus than they do near the nucleus of the hydrogen atoms, giving the atoms of oxygen and hydrogen slightly negative and positive charges, respectively. Another way of stating this is that the probability of finding a shared

electron near an oxygen nucleus is more likely than finding it near a hydrogen nucleus. Either way, the atom's relative electronegativity contributes to the development of partial charges whenever one element is significantly more electronegative than the other, and the charges generated by these polar bonds may then be used for the formation of hydrogen bonds based on the attraction of opposite partial charges. (Hydrogen bonds, which are discussed in detail below, are weak bonds between slightly positively charged hydrogen atoms to slightly negatively charged atoms in other molecules.) Since macromolecules often have atoms within them that differ in electronegativity, polar bonds are often present in organic molecules.

Nonpolar Covalent Bonds

Nonpolar covalent bonds form between two atoms of the same element or between different elements that share electrons equally. For example, molecular oxygen (O_2) is nonpolar because the electrons will be equally distributed between the two oxygen atoms.

Another example of a nonpolar covalent bond is methane (CH_4), also shown in [\[link\]](#). Carbon has four electrons in its outermost shell and needs four more to fill it. It gets these four from four hydrogen atoms, each atom providing one, making a stable outer shell of eight electrons. Carbon and hydrogen do not have the same electronegativity but are similar; thus, nonpolar bonds form. The hydrogen atoms each need one electron for their outermost shell, which is filled when it contains two electrons. These elements share the electrons equally among the carbons and the hydrogen atoms, creating a nonpolar covalent molecule.

	Bond type	Molecular shape	Molecular type
Water	 Polar covalent	 Bent	Polar
Methane	 Nonpolar covalent	 Tetrahedral	Nonpolar
Carbon dioxide	 Polar covalent	 Linear	Nonpolar

Whether a molecule is polar or nonpolar depends both on bond type and molecular shape. Both water and carbon dioxide have polar covalent bonds, but carbon dioxide is linear, so the partial charges on the molecule cancel each other out.

Hydrogen Bonds and Van Der Waals Interactions

Ionic and covalent bonds between elements require energy to break. Ionic bonds are not as strong as covalent, which determines their behavior in biological systems. However, not all bonds are ionic or covalent bonds. Weaker bonds can also form between molecules. Two weak bonds that occur frequently are hydrogen bonds and van der Waals interactions. Without these two types of bonds, life as we know it would not exist. Hydrogen bonds provide many of the critical, life-sustaining properties of

water and also stabilize the structures of proteins and DNA, the building block of cells.

When polar covalent bonds containing hydrogen form, the hydrogen in that bond has a slightly positive charge because hydrogen's electron is pulled more strongly toward the other element and away from the hydrogen. Because the hydrogen is slightly positive, it will be attracted to neighboring negative charges. When this happens, a weak interaction occurs between the δ^+ of the hydrogen from one molecule and the δ^- charge on the more electronegative atoms of another molecule, usually oxygen or nitrogen, or within the same molecule. This interaction is called a **hydrogen bond**. This type of bond is common and occurs regularly between water molecules. Individual hydrogen bonds are weak and easily broken; however, they occur in very large numbers in water and in organic polymers, creating a major force in combination. Hydrogen bonds are also responsible for zipping together the DNA double helix.

Like hydrogen bonds, **van der Waals interactions** are weak attractions or interactions between molecules. Van der Waals attractions can occur between any two or more molecules and are dependent on slight fluctuations of the electron densities, which are not always symmetrical around an atom. For these attractions to happen, the molecules need to be very close to one another. These bonds—along with ionic, covalent, and hydrogen bonds—contribute to the three-dimensional structure of the proteins in our cells that is necessary for their proper function.

Note:**Career Connection****Pharmaceutical Chemist**

Pharmaceutical chemists are responsible for the development of new drugs and trying to determine the mode of action of both old and new drugs. They are involved in every step of the drug development process. Drugs can be found in the natural environment or can be synthesized in the laboratory. In many cases, potential drugs found in nature are changed chemically in the laboratory to make them safer and more effective, and

sometimes synthetic versions of drugs substitute for the version found in nature.

After the initial discovery or synthesis of a drug, the chemist then develops the drug, perhaps chemically altering it, testing it to see if the drug is toxic, and then designing methods for efficient large-scale production. Then, the process of getting the drug approved for human use begins. In the United States, drug approval is handled by the Food and Drug Administration (FDA) and involves a series of large-scale experiments using human subjects to make sure the drug is not harmful and effectively treats the condition it aims to treat. This process often takes several years and requires the participation of physicians and scientists, in addition to chemists, to complete testing and gain approval.

An example of a drug that was originally discovered in a living organism is Paclitaxel (Taxol), an anti-cancer drug used to treat breast cancer. This drug was discovered in the bark of the pacific yew tree. Another example is aspirin, originally isolated from willow tree bark. Finding drugs often means testing hundreds of samples of plants, fungi, and other forms of life to see if any biologically active compounds are found within them.

Sometimes, traditional medicine can give modern medicine clues to where an active compound can be found. For example, the use of willow bark to make medicine has been known for thousands of years, dating back to ancient Egypt. It was not until the late 1800s, however, that the aspirin molecule, known as acetylsalicylic acid, was purified and marketed for human use.

Occasionally, drugs developed for one use are found to have unforeseen effects that allow these drugs to be used in other, unrelated ways. For example, the drug minoxidil (Rogaine) was originally developed to treat high blood pressure. When tested on humans, it was noticed that individuals taking the drug would grow new hair. Eventually the drug was marketed to men and women with baldness to restore lost hair.

The career of the pharmaceutical chemist may involve detective work, experimentation, and drug development, all with the goal of making human beings healthier.

Section Summary

Matter is anything that occupies space and has mass. It is made up of elements. All of the 92 elements that occur naturally have unique qualities that allow them to combine in various ways to create molecules, which in turn combine to form cells, tissues, organ systems, and organisms. Atoms, which consist of protons, neutrons, and electrons, are the smallest units of an element that retain all of the properties of that element. Electrons can be transferred, shared, or cause charge disparities between atoms to create bonds, including ionic, covalent, and hydrogen bonds, as well as van der Waals interactions.

Art Connections

Exercise:

Problem:

[\[link\]](#) How many neutrons do carbon-12 and carbon-13 have, respectively?

Solution:

[\[link\]](#) Carbon-12 has six neutrons. Carbon-13 has seven neutrons.

Exercise:

Problem:

[\[link\]](#) An atom may give, take, or share electrons with another atom to achieve a full valence shell, the most stable electron configuration. Looking at this figure, how many electrons do elements in group 1 need to lose in order to achieve a stable electron configuration? How many electrons do elements in groups 14 and 17 need to gain to achieve a stable configuration?

Solution:

[\[link\]](#) Elements in group 1 need to lose one electron to achieve a stable electron configuration. Elements in groups 14 and 17 need to gain four and one electrons, respectively, to achieve a stable configuration.

Review Questions

Exercise:

Problem:

If xenon has an atomic number of 54 and a mass number of 108, how many neutrons does it have?

- a. 54
- b. 27
- c. 100
- d. 108

Solution:

A

Exercise:

Problem:

Atoms that vary in the number of neutrons found in their nuclei are called _____.

- a. ions
- b. neutrons
- c. neutral atoms
- d. isotopes

Solution:

D

Exercise:

Problem:

Potassium has an atomic number of 19. What is its electron configuration?

- a. shells 1 and 2 are full, and shell 3 has nine electrons
- b. shells 1, 2 and 3 are full and shell 4 has three electrons
- c. shells 1, 2 and 3 are full and shell 4 has one electron
- d. shells 1, 2 and 3 are full and no other electrons are present

Solution:

C

Exercise:

Problem: Which type of bond represents a weak chemical bond?

- a. hydrogen bond
- b. atomic bond
- c. covalent bond
- d. nonpolar covalent bond

Solution:

A

Free Response**Exercise:**

Problem: What makes ionic bonds different from covalent bonds?

Solution:

Ionic bonds are created between ions. The electrons are not shared between the atoms, but rather are associated more with one ion than the other. Ionic bonds are strong bonds, but are weaker than covalent bonds, meaning it takes less energy to break an ionic bond compared with a covalent one.

Exercise:

Problem:

Why are hydrogen bonds and van der Waals interactions necessary for cells?

Solution:

Hydrogen bonds and van der Waals interactions form weak associations between different molecules or within different regions of the same molecule. They provide the structure and shape necessary for proteins and DNA within cells so that they function properly.

Glossary

anion

negative ion that is formed by an atom gaining one or more electrons

atom

the smallest unit of matter that retains all of the chemical properties of an element

atomic mass

calculated mean of the mass number for an element's isotopes

atomic number

total number of protons in an atom

balanced chemical equation

statement of a chemical reaction with the number of each type of atom equalized for both the products and reactants

cation

positive ion that is formed by an atom losing one or more electrons

chemical bond

interaction between two or more of the same or different atoms that results in the formation of molecules

chemical reaction

process leading to the rearrangement of atoms in molecules

chemical reactivity

the ability to combine and to chemically bond with each other

compound

substance composed of molecules consisting of atoms of at least two different elements

covalent bond

type of strong bond formed between two of the same or different elements; forms when electrons are shared between atoms

electrolyte

ion necessary for nerve impulse conduction, muscle contractions and water balance

electron

negatively charged subatomic particle that resides outside of the nucleus in the electron orbital; lacks functional mass and has a negative charge of -1 unit

electron configuration

arrangement of electrons in an atom's electron shell (for example, $1s^2 2s^2 2p^6$)

electron orbital

how electrons are spatially distributed surrounding the nucleus; the area where an electron is most likely to be found

electron transfer

movement of electrons from one element to another; important in creation of ionic bonds

electronegativity

ability of some elements to attract electrons (often of hydrogen atoms), acquiring partial negative charges in molecules and creating partial positive charges on the hydrogen atoms

element

one of 118 unique substances that cannot be broken down into smaller substances; each element has unique properties and a specified number of protons

equilibrium

steady state of relative reactant and product concentration in reversible chemical reactions in a closed system

hydrogen bond

weak bond between slightly positively charged hydrogen atoms to slightly negatively charged atoms in other molecules

inert gas

(also, noble gas) element with filled outer electron shell that is unreactive with other atoms

ion

atom or chemical group that does not contain equal numbers of protons and electrons

ionic bond

chemical bond that forms between ions with opposite charges (cations and anions)

irreversible chemical reaction

chemical reaction where reactants proceed uni-directionally to form products

isotope

one or more forms of an element that have different numbers of neutrons

law of mass action

chemical law stating that the rate of a reaction is proportional to the concentration of the reacting substances

mass number

total number of protons and neutrons in an atom

matter

anything that has mass and occupies space

molecule

two or more atoms chemically bonded together

neutron

uncharged particle that resides in the nucleus of an atom; has a mass of one amu

noble gas

see inert gas

nonpolar covalent bond

type of covalent bond that forms between atoms when electrons are shared equally between them

nucleus

core of an atom; contains protons and neutrons

octet rule

rule that atoms are most stable when they hold eight electrons in their outermost shells

orbital

region surrounding the nucleus; contains electrons

periodic table

organizational chart of elements indicating the atomic number and atomic mass of each element; provides key information about the properties of the elements

polar covalent bond

type of covalent bond that forms as a result of unequal sharing of electrons, resulting in the creation of slightly positive and slightly negative charged regions of the molecule

product

molecule found on the right side of a chemical equation

proton

positively charged particle that resides in the nucleus of an atom; has a mass of one amu and a charge of +1

radioisotope

isotope that emits radiation composed of subatomic particles to form more stable elements

reactant

molecule found on the left side of a chemical equation

reversible chemical reaction

chemical reaction that functions bi-directionally, where products may turn into reactants if their concentration is great enough

valence shell

outermost shell of an atom

van der Waals interaction

very weak interaction between molecules due to temporary charges attracting atoms that are very close together

Water

By the end of this section, you will be able to:

- Describe the properties of water that are critical to maintaining life
- Explain why water is an excellent solvent
- Provide examples of water's cohesive and adhesive properties
- Discuss the role of acids, bases, and buffers in homeostasis

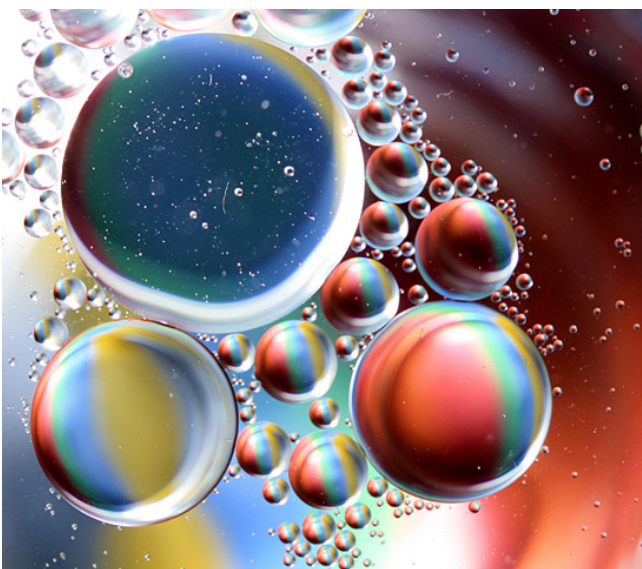
Why do scientists spend time looking for water on other planets? Why is water so important? It is because water is essential to life as we know it. Water is one of the more abundant molecules and the one most critical to life on Earth. Approximately 60–70 percent of the human body is made up of water. Without it, life as we know it simply would not exist.

The polarity of the water molecule and its resulting hydrogen bonding make water a unique substance with special properties that are intimately tied to the processes of life. Life originally evolved in a watery environment, and most of an organism's cellular chemistry and metabolism occur inside the watery contents of the cell's cytoplasm. Special properties of water are its high heat capacity and heat of vaporization, its ability to dissolve polar molecules, its cohesive and adhesive properties, and its dissociation into ions that leads to the generation of pH. Understanding these characteristics of water helps to elucidate its importance in maintaining life.

Water's Polarity

One of water's important properties is that it is composed of polar molecules: the hydrogen and oxygen within water molecules (H_2O) form polar covalent bonds. While there is no net charge to a water molecule, the polarity of water creates a slightly positive charge on hydrogen and a slightly negative charge on oxygen, contributing to water's properties of attraction. Water's charges are generated because oxygen is more electronegative than hydrogen, making it more likely that a shared electron would be found near the oxygen nucleus than the hydrogen nucleus, thus generating the partial negative charge near the oxygen.

As a result of water's polarity, each water molecule attracts other water molecules because of the opposite charges between water molecules, forming hydrogen bonds. Water also attracts or is attracted to other polar molecules and ions. A polar substance that interacts readily with or dissolves in water is referred to as **hydrophilic** (hydro- = “water”; -philic = “loving”). In contrast, non-polar molecules such as oils and fats do not interact well with water, as shown in [\[link\]](#) and separate from it rather than dissolve in it, as we see in salad dressings containing oil and vinegar (an acidic water solution). These nonpolar compounds are called **hydrophobic** (hydro- = “water”; -phobic = “fearing”).



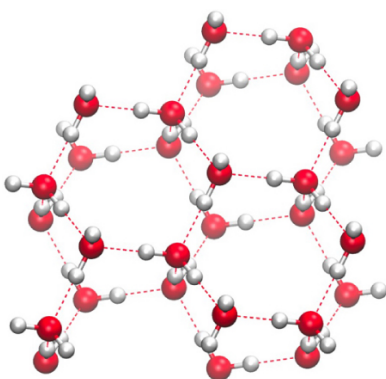
Oil and water do not mix. As this macro image of oil and water shows, oil does not dissolve in water but forms droplets instead. This is due to it being a nonpolar compound. (credit: Gautam Dogra).

Water's States: Gas, Liquid, and Solid

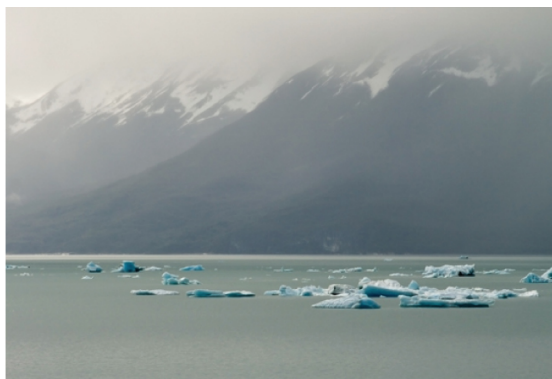
The formation of hydrogen bonds is an important quality of the liquid water that is crucial to life as we know it. As water molecules make hydrogen bonds with each other, water takes on some unique chemical characteristics compared to other liquids and, since living things have a high water content, understanding these chemical features is key to understanding life. In liquid water, hydrogen bonds are constantly formed and broken as the water molecules slide past each other. The breaking of these bonds is caused by the motion (kinetic energy) of the water molecules due to the heat contained in the system. When the heat is raised as water is boiled, the higher kinetic energy of the water molecules causes the hydrogen bonds to break completely and allows water molecules to escape into the air as gas (steam or water vapor). On the other hand, when the temperature of water is reduced and water freezes, the water molecules form a crystalline structure maintained by hydrogen bonding (there is not enough energy to break the hydrogen bonds) that makes ice less dense than liquid water, a phenomenon not seen in the solidification of other liquids.

Water's lower density in its solid form is due to the way hydrogen bonds are oriented as it freezes: the water molecules are pushed farther apart compared to liquid water. With most other liquids, solidification when the temperature drops includes the lowering of kinetic energy between molecules, allowing them to pack even more tightly than in liquid form and giving the solid a greater density than the liquid.

The lower density of ice, illustrated and pictured in [\[link\]](#), an anomaly, causes it to float at the surface of liquid water, such as in an iceberg or in the ice cubes in a glass of ice water. In lakes and ponds, ice will form on the surface of the water creating an insulating barrier that protects the animals and plant life in the pond from freezing. Without this layer of insulating ice, plants and animals living in the pond would freeze in the solid block of ice and could not survive. The detrimental effect of freezing on living organisms is caused by the expansion of ice relative to liquid water. The ice crystals that form upon freezing rupture the delicate membranes essential for the function of living cells, irreversibly damaging them. Cells can only survive freezing if the water in them is temporarily replaced by another liquid like glycerol.



(a)



(b)

Hydrogen bonding makes ice less dense than liquid water. The (a) lattice structure of ice makes it less dense than the freely flowing molecules of liquid water, enabling it to (b) float on water. (credit a: modification of work by Jane Whitney, image created using Visual Molecular Dynamics (VMD) software^[footnote]; credit b: modification of work by Carlos Ponte)

W. Humphrey W., A. Dalke, and K. Schulten, “VMD —Visual Molecular Dynamics,” *Journal of Molecular Graphics* 14 (1996): 33-38.

Note:

Link to Learning



Click [here](#) to see a 3-D animation of the structure of an ice lattice. (Image credit: Jane Whitney. Image created using Visual Molecular Dynamics

VMD software.^[footnote])

W. Humphrey W., A. Dalke, and K. Schulten, “VMD—Visual Molecular Dynamics,” *Journal of Molecular Graphics* 14 (1996): 33-38.

Water’s High Heat Capacity

Water’s high heat capacity is a property caused by hydrogen bonding among water molecules. Water has the highest **specific heat capacity** of any liquids. Specific heat is defined as the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius. For water, this amount is one **calorie**. It therefore takes water a long time to heat and long time to cool. In fact, the specific heat capacity of water is about five times more than that of sand. This explains why the land cools faster than the sea. Due to its high heat capacity, water is used by warm blooded animals to more evenly disperse heat in their bodies: it acts in a similar manner to a car’s cooling system, transporting heat from warm places to cool places, causing the body to maintain a more even temperature.

Water’s Heat of Vaporization

Water also has a high **heat of vaporization**, the amount of energy required to change one gram of a liquid substance to a gas. A considerable amount of heat energy (586 cal) is required to accomplish this change in water. This process occurs on the surface of water. As liquid water heats up, hydrogen bonding makes it difficult to separate the liquid water molecules from each other, which is required for it to enter its gaseous phase (steam). As a result, water acts as a heat sink or heat reservoir and requires much more heat to boil than does a liquid such as ethanol (grain alcohol), whose hydrogen bonding with other ethanol molecules is weaker than water’s hydrogen bonding. Eventually, as water reaches its boiling point of 100° Celsius (212° Fahrenheit), the heat is able to break the hydrogen bonds between the water molecules, and the kinetic energy (motion) between the water molecules allows them to escape from the liquid as a gas. Even when below its boiling point, water’s individual molecules acquire enough energy from

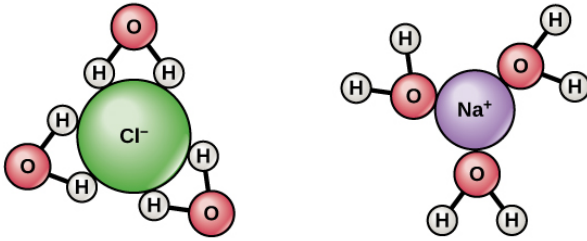
other water molecules such that some surface water molecules can escape and vaporize: this process is known as **evaporation**.

The fact that hydrogen bonds need to be broken for water to evaporate means that a substantial amount of energy is used in the process. As the water evaporates, energy is taken up by the process, cooling the environment where the evaporation is taking place. In many living organisms, including in humans, the evaporation of sweat, which is 90 percent water, allows the organism to cool so that homeostasis of body temperature can be maintained.

Water's Solvent Properties

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Therefore, water is referred to as a **solvent**, a substance capable of dissolving other polar molecules and ionic compounds. The charges associated with these molecules will form hydrogen bonds with water, surrounding the particle with water molecules. This is referred to as a **sphere of hydration**, or a hydration shell, as illustrated in [\[link\]](#) and serves to keep the particles separated or dispersed in the water.

When ionic compounds are added to water, the individual ions react with the polar regions of the water molecules and their ionic bonds are disrupted in the process of **dissociation**. Dissociation occurs when atoms or groups of atoms break off from molecules and form ions. Consider table salt (NaCl, or sodium chloride): when NaCl crystals are added to water, the molecules of NaCl dissociate into Na^+ and Cl^- ions, and spheres of hydration form around the ions, illustrated in [\[link\]](#). The positively charged sodium ion is surrounded by the partially negative charge of the water molecule's oxygen. The negatively charged chloride ion is surrounded by the partially positive charge of the hydrogen on the water molecule.



When table salt (NaCl) is mixed in water, spheres of hydration are formed around the ions.

Water's Cohesive and Adhesive Properties

Have you ever filled a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water forms a dome-like shape above the rim of the glass. This water can stay above the glass because of the property of **cohesion**. In cohesion, water molecules are attracted to each other (because of hydrogen bonding), keeping the molecules together at the liquid-gas (water-air) interface, although there is no more room in the glass.

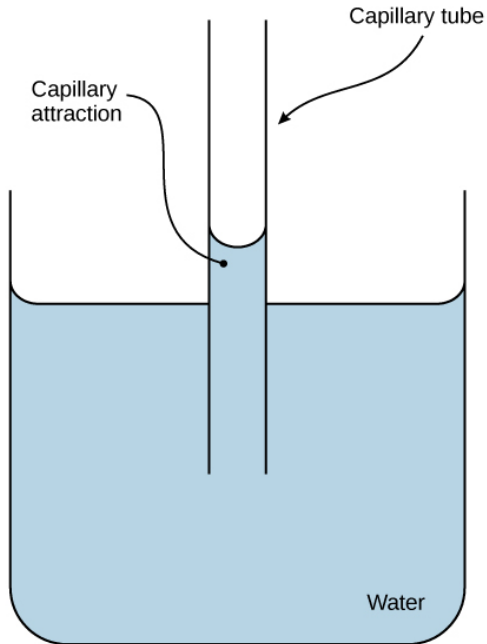
Cohesion allows for the development of **surface tension**, the capacity of a substance to withstand being ruptured when placed under tension or stress. This is also why water forms droplets when placed on a dry surface rather than being flattened out by gravity. When a small scrap of paper is placed onto the droplet of water, the paper floats on top of the water droplet even though paper is denser (heavier) than the water. Cohesion and surface tension keep the hydrogen bonds of water molecules intact and support the item floating on the top. It's even possible to "float" a needle on top of a glass of water if it is placed gently without breaking the surface tension, as shown in [\[link\]](#).



The weight of the needle is pulling the surface downward; at the same time, the surface tension is pulling it up, suspending it on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle.

(credit: Cory Zanker)

These cohesive forces are related to water's property of **adhesion**, or the attraction between water molecules and other molecules. This attraction is sometimes stronger than water's cohesive forces, especially when the water is exposed to charged surfaces such as those found on the inside of thin glass tubes known as capillary tubes. Adhesion is observed when water "climbs" up the tube placed in a glass of water: notice that the water appears to be higher on the sides of the tube than in the middle. This is because the water molecules are attracted to the charged glass walls of the capillary more than they are to each other and therefore adhere to it. This type of adhesion is called **capillary action**, and is illustrated in [\[link\]](#).



Capillary action in a glass tube is caused by the adhesive forces exerted by the internal surface of the glass exceeding the cohesive forces between the water molecules themselves. (credit: modification of work by Pearson-Scott Foresman, donated to the Wikimedia Foundation)

Why are cohesive and adhesive forces important for life? Cohesive and adhesive forces are important for the transport of water from the roots to the leaves in plants. These forces create a “pull” on the water column. This pull results from the tendency of water molecules being evaporated on the surface of the plant to stay connected to water molecules below them, and so they are pulled along. Plants use this natural phenomenon to help transport water from their roots to their leaves. Without these properties of water, plants would be unable to receive the water and the dissolved minerals they require. In another example, insects such as the water strider,

shown in [\[link\]](#), use the surface tension of water to stay afloat on the surface layer of water and even mate there.

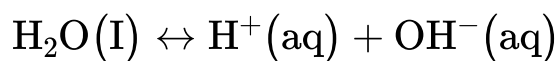


Water's cohesive and adhesive properties allow this water strider (*Gerris* sp.) to stay afloat. (credit: Tim Vickers)

pH, Buffers, Acids, and Bases

The pH of a solution indicates its acidity or alkalinity.

Equation:



litmus or pH paper, filter paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator, to test how much acid (acidity) or base (alkalinity) exists in a solution. You might have even used some to test whether the water in a swimming pool is properly treated. In both cases, the pH test measures the concentration of hydrogen ions in a given solution.

Hydrogen ions are spontaneously generated in pure water by the dissociation (ionization) of a small percentage of water molecules into

equal numbers of hydrogen (H^+) ions and hydroxide (OH^-) ions. While the hydroxide ions are kept in solution by their hydrogen bonding with other water molecules, the hydrogen ions, consisting of naked protons, are immediately attracted to un-ionized water molecules, forming hydronium ions (H_3O^+). Still, by convention, scientists refer to hydrogen ions and their concentration as if they were free in this state in liquid water.

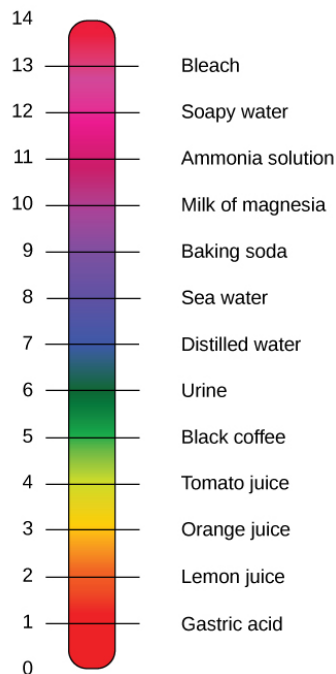
The concentration of hydrogen ions dissociating from pure water is 1×10^{-7} moles H^+ ions per liter of water. Moles (mol) are a way to express the amount of a substance (which can be atoms, molecules, ions, etc), with one mole being equal to 6.02×10^{23} particles of the substance. Therefore, 1 mole of water is equal to 6.02×10^{23} water molecules. The pH is calculated as the negative of the base 10 logarithm of this concentration. The \log_{10} of 1×10^{-7} is -7.0, and the negative of this number (indicated by the “p” of “pH”) yields a pH of 7.0, which is also known as neutral pH. The pH inside of human cells and blood are examples of two areas of the body where near-neutral pH is maintained.

Non-neutral pH readings result from dissolving acids or bases in water. Using the negative logarithm to generate positive integers, high concentrations of hydrogen ions yield a low pH number, whereas low levels of hydrogen ions result in a high pH. An **acid** is a substance that increases the concentration of hydrogen ions (H^+) in a solution, usually by having one of its hydrogen atoms dissociate. A **base** provides either hydroxide ions (OH^-) or other negatively charged ions that combine with hydrogen ions, reducing their concentration in the solution and thereby raising the pH. In cases where the base releases hydroxide ions, these ions bind to free hydrogen ions, generating new water molecules.

The stronger the acid, the more readily it donates H^+ . For example, hydrochloric acid (HCl) completely dissociates into hydrogen and chloride ions and is highly acidic, whereas the acids in tomato juice or vinegar do not completely dissociate and are considered weak acids. Conversely, strong bases are those substances that readily donate OH^- or take up hydrogen ions. Sodium hydroxide (NaOH) and many household cleaners are highly alkaline and give up OH^- rapidly when placed in water, thereby raising the pH. An example of a weak basic solution is seawater, which has

a pH near 8.0, close enough to neutral pH that marine organisms adapted to this saline environment are able to thrive in it.

The **pH scale** is, as previously mentioned, an inverse logarithm and ranges from 0 to 14 ([link](#)). Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. Extremes in pH in either direction from 7.0 are usually considered inhospitable to life. The pH inside cells (6.8) and the pH in the blood (7.4) are both very close to neutral. However, the environment in the stomach is highly acidic, with a pH of 1 to 2. So how do the cells of the stomach survive in such an acidic environment? How do they homeostatically maintain the near neutral pH inside them? The answer is that they cannot do it and are constantly dying. New stomach cells are constantly produced to replace dead ones, which are digested by the stomach acids. It is estimated that the lining of the human stomach is completely replaced every seven to ten days.



The pH scale measures the concentration of hydrogen ions (H^+) in a solution.

(credit: modification of work
by Edward Stevens)

Note:

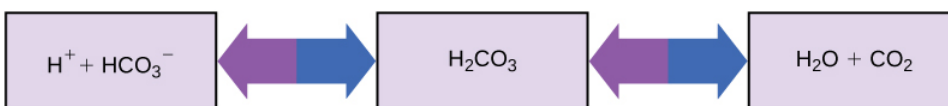
Link to Learning



Watch this video for a straightforward explanation of pH and its logarithmic scale.

https://www.openstaxcollege.org/l/pH_scale

So how can organisms whose bodies require a near-neutral pH ingest acidic and basic substances (a human drinking orange juice, for example) and survive? Buffers are the key. **Buffers** readily absorb excess H^+ or OH^- , keeping the pH of the body carefully maintained in the narrow range required for survival. Maintaining a constant blood pH is critical to a person's well-being. The buffer maintaining the pH of human blood involves carbonic acid (H_2CO_3), bicarbonate ion (HCO_3^-), and carbon dioxide (CO_2). When bicarbonate ions combine with free hydrogen ions and become carbonic acid, hydrogen ions are removed, moderating pH changes. Similarly, as shown in [\[link\]](#), excess carbonic acid can be converted to carbon dioxide gas and exhaled through the lungs. This prevents too many free hydrogen ions from building up in the blood and dangerously reducing the blood's pH. Likewise, if too much OH^- is introduced into the system, carbonic acid will combine with it to create bicarbonate, lowering the pH. Without this buffer system, the body's pH would fluctuate enough to put survival in jeopardy.



This diagram shows the body's buffering of blood pH levels. The blue arrows show the process of raising pH as more CO_2 is made. The purple arrows indicate the reverse process: the lowering of pH as more bicarbonate is created.

Other examples of buffers are antacids used to combat excess stomach acid. Many of these over-the-counter medications work in the same way as blood buffers, usually with at least one ion capable of absorbing hydrogen and moderating pH, bringing relief to those that suffer “heartburn” after eating. The unique properties of water that contribute to this capacity to balance pH—as well as water’s other characteristics—are essential to sustaining life on Earth.

Note:

Link to Learning



To learn more about water. Visit the [U.S. Geological Survey Water Science for Schools](https://www.usgs.gov/edu/water-science-for-schools) All About Water! website.

Section Summary

Water has many properties that are critical to maintaining life. It is a polar molecule, allowing for the formation of hydrogen bonds. Hydrogen bonds allow ions and other polar molecules to dissolve in water. Therefore, water is an excellent solvent. The hydrogen bonds between water molecules cause the water to have a high heat capacity, meaning it takes a lot of added heat to raise its temperature. As the temperature rises, the hydrogen bonds between water continually break and form anew. This allows for the overall temperature to remain stable, although energy is added to the system. Water also exhibits a high heat of vaporization, which is key to how organisms cool themselves by the evaporation of sweat. Water's cohesive forces allow for the property of surface tension, whereas its adhesive properties are seen as water rises inside capillary tubes. The pH value is a measure of hydrogen ion concentration in a solution and is one of many chemical characteristics that is highly regulated in living organisms through homeostasis. Acids and bases can change pH values, but buffers tend to moderate the changes they cause. These properties of water are intimately connected to the biochemical and physical processes performed by living organisms, and life would be very different if these properties were altered, if it could exist at all.

Review Questions

Exercise:

Problem: Which of the following statements is not true?

- a. Water is polar.
- b. Water stabilizes temperature.
- c. Water is essential for life.
- d. Water is the most abundant molecule in the Earth's atmosphere.

Solution:

D

Exercise:

Problem:

When acids are added to a solution, the pH should _____.

- a. decrease
- b. increase
- c. stay the same
- d. cannot tell without testing

Solution:

A

Exercise:**Problem:**

A molecule that binds up excess hydrogen ions in a solution is called a(n) _____.

- a. acid
- b. isotope
- c. base
- d. donator

Solution:

C

Exercise:

Problem: Which of the following statements is true?

- a. Acids and bases cannot mix together.
- b. Acids and bases will neutralize each other.
- c. Acids, but not bases, can change the pH of a solution.

d. Acids donate hydroxide ions (OH^-); bases donate hydrogen ions (H^+).

Solution:

B

Free Response

Exercise:

Problem: Discuss how buffers help prevent drastic swings in pH.

Solution:

Buffers absorb the free hydrogen ions and hydroxide ions that result from chemical reactions. Because they can bond these ions, they prevent increases or decreases in pH. An example of a buffer system is the bicarbonate system in the human body. This system is able to absorb hydrogen and hydroxide ions to prevent changes in pH and keep cells functioning properly.

Exercise:

Problem: Why can some insects walk on water?

Solution:

Some insects can walk on water, although they are heavier (denser) than water, because of the surface tension of water. Surface tension results from cohesion, or the attraction between water molecules at the surface of the body of water (the liquid-air/gas interface).

Glossary

acid

molecule that donates hydrogen ions and increases the concentration of hydrogen ions in a solution

adhesion

attraction between water molecules and other molecules

base

molecule that donates hydroxide ions or otherwise binds excess hydrogen ions and decreases the concentration of hydrogen ions in a solution

buffer

substance that prevents a change in pH by absorbing or releasing hydrogen or hydroxide ions

calorie

amount of heat required to change the temperature of one gram of water by one degree Celsius

capillary action

occurs because water molecules are attracted to charges on the inner surfaces of narrow tubular structures such as glass tubes, drawing the water molecules to the sides of the tubes

cohesion

intermolecular forces between water molecules caused by the polar nature of water; responsible for surface tension

dissociation

release of an ion from a molecule such that the original molecule now consists of an ion and the charged remains of the original, such as when water dissociates into H^+ and OH^-

evaporation

separation of individual molecules from the surface of a body of water, leaves of a plant, or the skin of an organism

heat of vaporization of water

high amount of energy required for liquid water to turn into water vapor

hydrophilic

describes ions or polar molecules that interact well with other polar molecules such as water

hydrophobic

describes uncharged non-polar molecules that do not interact well with polar molecules such as water

litmus paper

(also, pH paper) filter paper that has been treated with a natural water-soluble dye that changes its color as the pH of the environment changes so it can be used as a pH indicator

pH paper

see litmus paper

pH scale

scale ranging from zero to 14 that is inversely proportional to the concentration of hydrogen ions in a solution

solvent

substance capable of dissolving another substance

specific heat capacity

the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius

sphere of hydration

when a polar water molecule surrounds charged or polar molecules thus keeping them dissolved and in solution

surface tension

tension at the surface of a body of liquid that prevents the molecules from separating; created by the attractive cohesive forces between the

molecules of the liquid

Carbon

By the end of this section, you will be able to:

- Explain why carbon is important for life
- Describe the role of functional groups in biological molecules

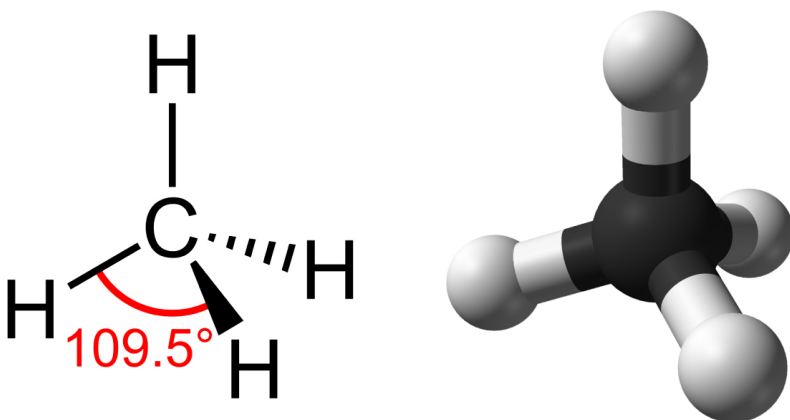
Cells are made of many complex molecules called macromolecules, such as proteins, nucleic acids (RNA and DNA), carbohydrates, and lipids. The macromolecules are a subset of **organic molecules** (any carbon-containing liquid, solid, or gas) that are especially important for life. The fundamental component for all of these macromolecules is carbon. The carbon atom has unique properties that allow it to form covalent bonds to as many as four different atoms, making this versatile element ideal to serve as the basic structural component, or “backbone,” of the macromolecules.

Individual carbon atoms have an incomplete outermost electron shell. With an atomic number of 6 (six electrons and six protons), the first two electrons fill the inner shell, leaving four in the second shell. Therefore, carbon atoms can form up to four covalent bonds with other atoms to satisfy the octet rule. The methane molecule provides an example: it has the chemical formula CH_4 . Each of its four hydrogen atoms forms a single covalent bond with the carbon atom by sharing a pair of electrons. This results in a filled outermost shell.

Hydrocarbons

Hydrocarbons are organic molecules consisting entirely of carbon and hydrogen, such as methane (CH_4) described above. We often use hydrocarbons in our daily lives as fuels—like the propane in a gas grill or the butane in a lighter. The many covalent bonds between the atoms in hydrocarbons store a great amount of energy, which is released when these molecules are burned (oxidized). Methane, an excellent fuel, is the simplest hydrocarbon molecule, with a central carbon atom bonded to four different hydrogen atoms, as illustrated in [\[link\]](#). The geometry of the methane molecule, where the atoms reside in three dimensions, is determined by the shape of its electron orbitals. The carbons and the four hydrogen atoms

form a shape known as a tetrahedron, with four triangular faces; for this reason, methane is described as having tetrahedral geometry.



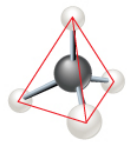
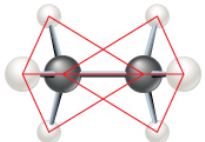
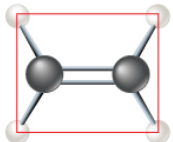
Methane has a tetrahedral geometry, with each of the four hydrogen atoms spaced 109.5° apart.

As the backbone of the large molecules of living things, hydrocarbons may exist as linear carbon chains, carbon rings, or combinations of both. Furthermore, individual carbon-to-carbon bonds may be single, double, or triple covalent bonds, and each type of bond affects the geometry of the molecule in a specific way. This three-dimensional shape or conformation of the large molecules of life (macromolecules) is critical to how they function.

Hydrocarbon Chains

Hydrocarbon chains are formed by successive bonds between carbon atoms and may be branched or unbranched. Furthermore, the overall geometry of the molecule is altered by the different geometries of single, double, and triple covalent bonds, illustrated in [\[link\]](#). The hydrocarbons ethane, ethene, and ethyne serve as examples of how different carbon-to-carbon bonds

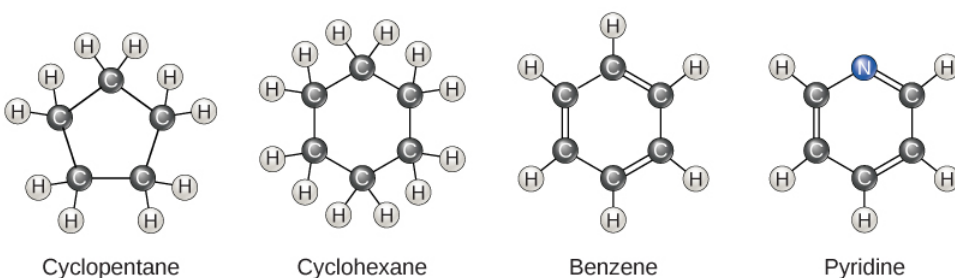
affect the geometry of the molecule. The names of all three molecules start with the prefix “eth-,” which is the prefix for two carbon hydrocarbons. The suffixes “-ane,” “-ene,” and “-yne” refer to the presence of single, double, or triple carbon-carbon bonds, respectively. Thus, propane, propene, and propyne follow the same pattern with three carbon molecules, butane, butene, and butyne for four carbon molecules, and so on. Double and triple bonds change the geometry of the molecule: single bonds allow rotation along the axis of the bond, whereas double bonds lead to a planar configuration and triple bonds to a linear one. These geometries have a significant impact on the shape a particular molecule can assume.

Methane (CH ₄)	Ethane (C ₂ H ₆)	Ethene (C ₂ H ₄)
		
Tetrahedral (single bond)	Tetrahedral (single bond)	Planar (double bond)

When carbon forms single bonds with other atoms, the shape is tetrahedral. When two carbon atoms form a double bond, the shape is planar, or flat. Single bonds, like those found in ethane, are able to rotate. Double bonds, like those found in ethene cannot rotate, so the atoms on either side are locked in place.

Hydrocarbon Rings

So far, the hydrocarbons we have discussed have been **aliphatic hydrocarbons**, which consist of linear chains of carbon atoms. Another type of hydrocarbon, **aromatic hydrocarbons**, consists of closed rings of carbon atoms. Ring structures are found in hydrocarbons, sometimes with the presence of double bonds, which can be seen by comparing the structure of cyclohexane to benzene in [\[link\]](#). Examples of biological molecules that incorporate the benzene ring include some amino acids and cholesterol and its derivatives, including the hormones estrogen and testosterone. The benzene ring is also found in the herbicide 2,4-D. Benzene is a natural component of crude oil and has been classified as a carcinogen. Some hydrocarbons have both aliphatic and aromatic portions; beta-carotene is an example of such a hydrocarbon.



Carbon can form five- and six-membered rings. Single or double bonds may connect the carbons in the ring, and nitrogen may be substituted for carbon.

Isomers

The three-dimensional placement of atoms and chemical bonds within organic molecules is central to understanding their chemistry. Molecules that share the same chemical formula but differ in the placement (structure) of their atoms and/or chemical bonds are known as **isomers**. **Structural isomers** (like butane and isobutene shown in [\[link\]](#) a) differ in the placement of their covalent bonds: both molecules have four carbons and ten hydrogens (C_4H_{10}), but the different arrangement of the atoms within

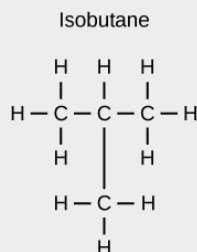
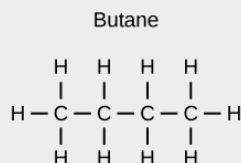
the molecules leads to differences in their chemical properties. For example, due to their different chemical properties, butane is suited for use as a fuel for cigarette lighters and torches, whereas isobutene is suited for use as a refrigerant and a propellant in spray cans.

Geometric isomers, on the other hand, have similar placements of their covalent bonds but differ in how these bonds are made to the surrounding atoms, especially in carbon-to-carbon double bonds. In the simple molecule butene (C_4H_8), the two methyl groups (CH_3) can be on either side of the double covalent bond central to the molecule, as illustrated in [\[link\]](#) **b**. When the carbons are bound on the same side of the double bond, this is the *cis* configuration; if they are on opposite sides of the double bond, it is a *trans* configuration. In the *trans* configuration, the carbons form a more or less linear structure, whereas the carbons in the *cis* configuration make a bend (change in direction) of the carbon backbone.

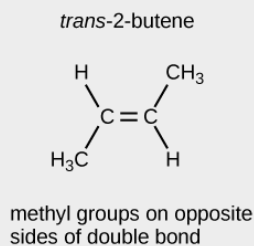
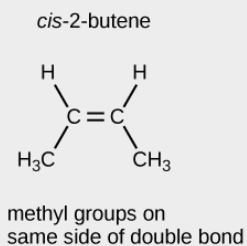
Note:

Art Connection

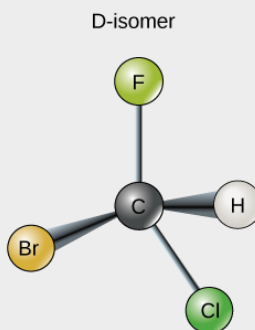
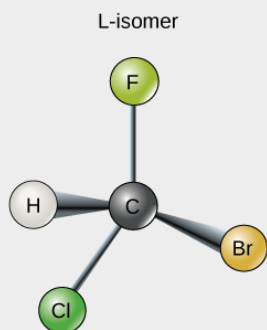
(a) Structural isomers



(b) Geometric isomers



(c) Enantiomers

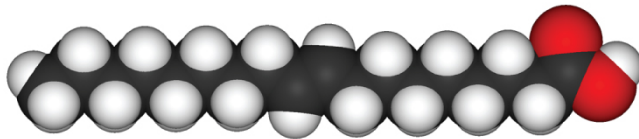


Molecules that have the same number and type of atoms arranged differently are called isomers. (a) Structural isomers have a different covalent arrangement of atoms. (b) Geometric isomers have a different arrangement of atoms around a double bond. (c) Enantiomers are mirror images of each other.

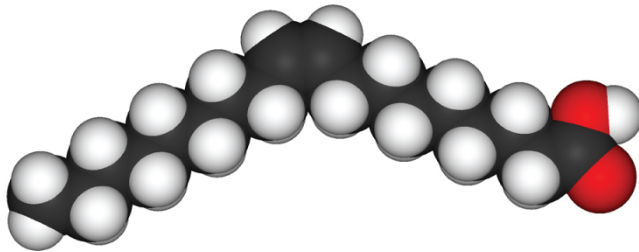
Which of the following statements is false?

- a. Molecules with the formulas $\text{CH}_3\text{CH}_2\text{COOH}$ and $\text{C}_3\text{H}_6\text{O}_2$ could be structural isomers.
- b. Molecules must have a double bond to be *cis-trans* isomers.
- c. To be enantiomers, a molecule must have at least three different atoms or groups connected to a central carbon.
- d. To be enantiomers, a molecule must have at least four different atoms or groups connected to a central carbon.

In triglycerides (fats and oils), long carbon chains known as fatty acids may contain double bonds, which can be in either the *cis* or *trans* configuration, illustrated in [\[link\]](#). Fats with at least one double bond between carbon atoms are unsaturated fats. When some of these bonds are in the *cis* configuration, the resulting bend in the carbon backbone of the chain means that triglyceride molecules cannot pack tightly, so they remain liquid (oil) at room temperature. On the other hand, triglycerides with *trans* double bonds (popularly called trans fats), have relatively linear fatty acids that are able to pack tightly together at room temperature and form solid fats. In the human diet, trans fats are linked to an increased risk of cardiovascular disease, so many food manufacturers have reduced or eliminated their use in recent years. In contrast to unsaturated fats, triglycerides without double bonds between carbon atoms are called saturated fats, meaning that they contain all the hydrogen atoms available. Saturated fats are a solid at room temperature and usually of animal origin.



Elaidic acid

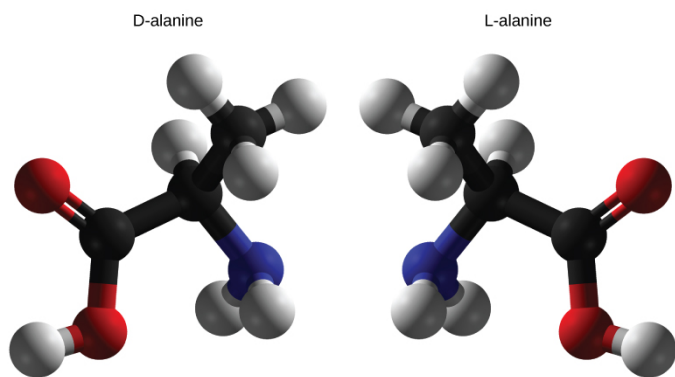


Oleic acid

These space-filling models show a *cis* (oleic acid) and a *trans* (elaidic acid) fatty acid. Notice the bend in the molecule cause by the *cis* configuration.

Enantiomers

Enantiomers are molecules that share the same chemical structure and chemical bonds but differ in the three-dimensional placement of atoms so that they are mirror images. As shown in [\[link\]](#), an amino acid alanine example, the two structures are non-superimposable. In nature, only the L-forms of amino acids are used to make proteins. Some D forms of amino acids are seen in the cell walls of bacteria, but never in their proteins. Similarly, the D-form of glucose is the main product of photosynthesis and the L-form of the molecule is rarely seen in nature.



D-alanine and L-alanine are examples of enantiomers or mirror images. Only the L-forms of amino acids are used to make proteins.

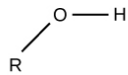
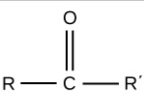
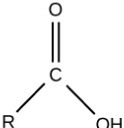
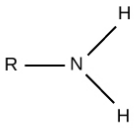
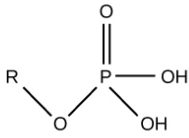
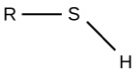
Functional Groups

Functional groups are groups of atoms that occur within molecules and confer specific chemical properties to those molecules. They are found along the “carbon backbone” of macromolecules. This carbon backbone is formed by chains and/or rings of carbon atoms with the occasional substitution of an element such as nitrogen or oxygen. Molecules with other elements in their carbon backbone are **substituted hydrocarbons**.

The functional groups in a macromolecule are usually attached to the carbon backbone at one or several different places along its chain and/or ring structure. Each of the four types of macromolecules—proteins, lipids, carbohydrates, and nucleic acids—has its own characteristic set of functional groups that contributes greatly to its differing chemical properties and its function in living organisms.

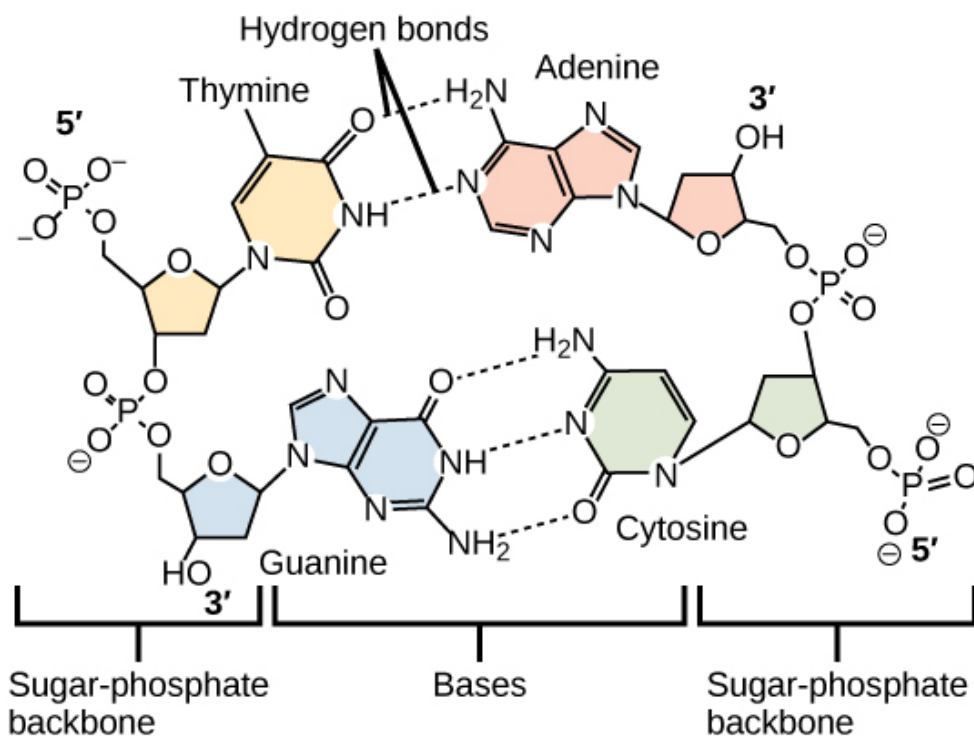
A functional group can participate in specific chemical reactions. Some of the important functional groups in biological molecules are shown in [\[link\]](#); they include: hydroxyl, methyl, carbonyl, carboxyl, amino, phosphate, and sulfhydryl. These groups play an important role in the formation of molecules like DNA, proteins, carbohydrates, and lipids. Functional groups

are usually classified as hydrophobic or hydrophilic depending on their charge or polarity characteristics. An example of a hydrophobic group is the non-polar methane molecule. Among the hydrophilic functional groups is the carboxyl group found in amino acids, some amino acid side chains, and the fatty acids that form triglycerides and phospholipids. This carboxyl group ionizes to release hydrogen ions (H^+) from the $COOH$ group resulting in the negatively charged COO^- group; this contributes to the hydrophilic nature of whatever molecule it is found on. Other functional groups, such as the carbonyl group, have a partially negatively charged oxygen atom that may form hydrogen bonds with water molecules, again making the molecule more hydrophilic.

Functional Group	Structure	Properties
Hydroxyl		Polar
Methyl	$R - CH_3$	Nonpolar
Carbonyl		Polar
Carboxyl		Charged, ionizes to release H^+ . Since carboxyl groups can release H^+ ions into solution, they are considered acidic.
Amino		Charged, accepts H^+ to form NH_3^+ . Since amino groups can remove H^+ from solution, they are considered basic.
Phosphate		Charged, ionizes to release H^+ . Since phosphate groups can release H^+ ions into solution, they are considered acidic.
Sulfhydryl		Polar

The functional groups shown here are found in many different biological molecules.

Hydrogen bonds between functional groups (within the same molecule or between different molecules) are important to the function of many macromolecules and help them to fold properly into and maintain the appropriate shape for functioning. Hydrogen bonds are also involved in various recognition processes, such as DNA complementary base pairing and the binding of an enzyme to its substrate, as illustrated in [\[link\]](#).



Hydrogen bonds connect two strands of DNA together to create the double-helix structure.

Section Summary

The unique properties of carbon make it a central part of biological molecules. Carbon binds to oxygen, hydrogen, and nitrogen covalently to

form the many molecules important for cellular function. Carbon has four electrons in its outermost shell and can form four bonds. Carbon and hydrogen can form hydrocarbon chains or rings. Functional groups are groups of atoms that confer specific properties to hydrocarbon (or substituted hydrocarbon) chains or rings that define their overall chemical characteristics and function.

Art Connections

Exercise:

Problem: [\[link\]](#) Which of the following statements is false?

- a. Molecules with the formulas $\text{CH}_3\text{CH}_2\text{COOH}$ and $\text{C}_3\text{H}_6\text{O}_2$ could be structural isomers.
- b. Molecules must have a double bond to be *cis-trans* isomers.
- c. To be enantiomers, a molecule must have at least three different atoms or groups connected to a central carbon.
- d. To be enantiomers, a molecule must have at least four different atoms or groups connected to a central carbon.

Solution:

[\[link\]](#) C

Review Questions

Exercise:

Problem:

Each carbon molecule can bond with as many as _____ other atom(s) or molecule(s).

- a. one
- b. two

- c. six
- d. four

Solution:

D

Exercise:

Problem:

Which of the following is not a functional group that can bond with carbon?

- a. sodium
- b. hydroxyl
- c. phosphate
- d. carbonyl

Solution:

A

Free Response

Exercise:

Problem: What property of carbon makes it essential for organic life?

Solution:

Carbon is unique and found in all living things because it can form up to four covalent bonds between atoms or molecules. These can be nonpolar or polar covalent bonds, and they allow for the formation of long chains of carbon molecules that combine to form proteins and DNA.

Exercise:**Problem:**

Compare and contrast saturated and unsaturated triglycerides.

Solution:

Saturated triglycerides contain no double bonds between carbon atoms; they are usually solid at room temperature. Unsaturated triglycerides contain at least one double bond between carbon atoms and are usually liquid at room temperature.

Glossary

aliphatic hydrocarbon

hydrocarbon consisting of a linear chain of carbon atoms

aromatic hydrocarbon

hydrocarbon consisting of closed rings of carbon atoms

enantiomers

molecules that share overall structure and bonding patterns, but differ in how the atoms are three dimensionally placed such that they are mirror images of each other

functional group

group of atoms that provides or imparts a specific function to a carbon skeleton

geometric isomer

isomer with similar bonding patterns differing in the placement of atoms alongside a double covalent bond

hydrocarbon

molecule that consists only of carbon and hydrogen

isomers

molecules that differ from one another even though they share the same chemical formula

organic molecule

any molecule containing carbon (except carbon dioxide)

structural isomers

molecules that share a chemical formula but differ in the placement of their chemical bonds

substituted hydrocarbon

hydrocarbon chain or ring containing an atom of another element in place of one of the backbone carbons

Introduction

class="introduction"

Foods such as bread, fruit, and cheese are rich sources of biological macromolecules . (credit: modification of work by Bengt Nyman)



Food provides the body with the nutrients it needs to survive. Many of these critical nutrients are biological macromolecules, or large molecules, necessary for life. These macromolecules (polymers) are built from

different combinations of smaller organic molecules (monomers). What specific types of biological macromolecules do living things require? How are these molecules formed? What functions do they serve? In this chapter, these questions will be explored.

Synthesis of Biological Macromolecules

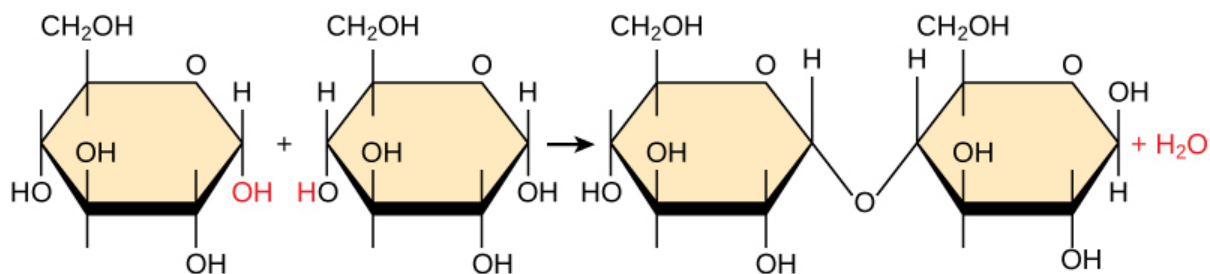
By the end of this section, you will be able to:

- Understand the synthesis of macromolecules
- Explain dehydration (or condensation) and hydrolysis reactions

As you've learned, **biological macromolecules** are large molecules, necessary for life, that are built from smaller organic molecules. There are four major classes of biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids); each is an important cell component and performs a wide array of functions. Combined, these molecules make up the majority of a cell's dry mass (recall that water makes up the majority of its complete mass). Biological macromolecules are organic, meaning they contain carbon. In addition, they may contain hydrogen, oxygen, nitrogen, and additional minor elements.

Dehydration Synthesis

Most macromolecules are made from single subunits, or building blocks, called **monomers**. The monomers combine with each other using covalent bonds to form larger molecules known as **polymers**. In doing so, monomers release water molecules as byproducts. This type of reaction is known as **dehydration synthesis**, which means “to put together while losing water.”

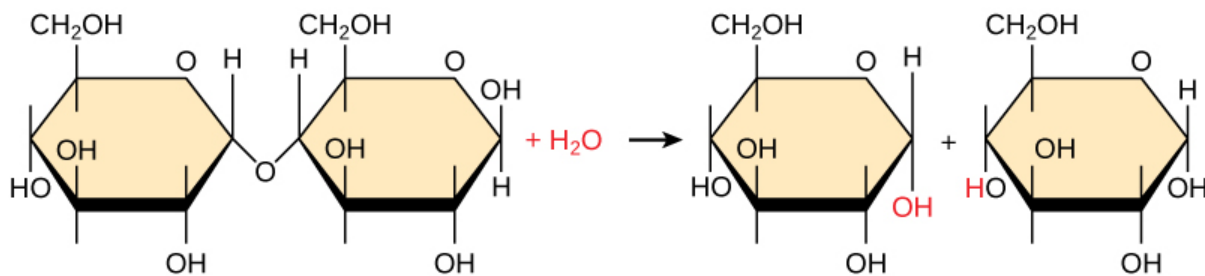


In the dehydration synthesis reaction depicted above, two molecules of glucose are linked together to form the disaccharide maltose. In the process, a water molecule is formed.

In a dehydration synthesis reaction ([\[link\]](#)), the hydrogen of one monomer combines with the hydroxyl group of another monomer, releasing a molecule of water. At the same time, the monomers share electrons and form covalent bonds. As additional monomers join, this chain of repeating monomers forms a polymer. Different types of monomers can combine in many configurations, giving rise to a diverse group of macromolecules. Even one kind of monomer can combine in a variety of ways to form several different polymers: for example, glucose monomers are the constituents of starch, glycogen, and cellulose.

Hydrolysis

Polymers are broken down into monomers in a process known as hydrolysis, which means “to split water,” a reaction in which a water molecule is used during the breakdown ([\[link\]](#)). During these reactions, the polymer is broken into two components: one part gains a hydrogen atom (H^+) and the other gains a hydroxyl molecule (OH^-) from a split water molecule.



In the hydrolysis reaction shown here, the disaccharide maltose is broken down to form two glucose monomers with the addition of a water molecule. Note that this reaction is the reverse of the synthesis reaction shown in [\[link\]](#).

Dehydration and **hydrolysis reactions** are catalyzed, or “sped up,” by specific enzymes; dehydration reactions involve the formation of new bonds, requiring energy, while hydrolysis reactions break bonds and release energy. These reactions are similar for most macromolecules, but each monomer and polymer reaction is specific for its class. For example, in our bodies, food is hydrolyzed, or broken down, into smaller molecules by catalytic enzymes in the digestive system. This allows for easy absorption of nutrients by cells in the intestine. Each macromolecule is broken down by a specific enzyme. For instance, carbohydrates are broken down by amylase, sucrase, lactase, or maltase. Proteins are broken down by the enzymes pepsin and peptidase, and by hydrochloric acid. Lipids are broken down by lipases. Breakdown of these macromolecules provides energy for cellular activities.

Note:

Link to Learning



Visit [this site](#) to see visual representations of dehydration synthesis and hydrolysis.

Section Summary

Proteins, carbohydrates, nucleic acids, and lipids are the four major classes of biological macromolecules—large molecules necessary for life that are built from smaller organic molecules. Macromolecules are made up of single units known as monomers that are joined by covalent bonds to form larger polymers. The polymer is more than the sum of its parts: it acquires

new characteristics, and leads to an osmotic pressure that is much lower than that formed by its ingredients; this is an important advantage in the maintenance of cellular osmotic conditions. A monomer joins with another monomer with the release of a water molecule, leading to the formation of a covalent bond. These types of reactions are known as dehydration or condensation reactions. When polymers are broken down into smaller units (monomers), a molecule of water is used for each bond broken by these reactions; such reactions are known as hydrolysis reactions. Dehydration and hydrolysis reactions are similar for all macromolecules, but each monomer and polymer reaction is specific to its class. Dehydration reactions typically require an investment of energy for new bond formation, while hydrolysis reactions typically release energy by breaking bonds.

Review Questions

Exercise:

Problem: Dehydration synthesis leads to formation of

- a. monomers
- b. polymers
- c. water and polymers
- d. none of the above

Solution:

C

Exercise:

Problem:

During the breakdown of polymers, which of the following reactions takes place?

- a. hydrolysis
- b. dehydration

- c. condensation
 - d. covalent bond
-

Solution:

A

Free Response

Exercise:

Problem: Why are biological macromolecules considered organic?

Solution:

Biological macromolecules are organic because they contain carbon.

Exercise:

Problem:

What role do electrons play in dehydration synthesis and hydrolysis?

Solution:

In a dehydration synthesis reaction, the hydrogen of one monomer combines with the hydroxyl group of another monomer, releasing a molecule of water. This creates an opening in the outer shells of atoms in the monomers, which can share electrons and form covalent bonds.

Glossary

biological macromolecule

large molecule necessary for life that is built from smaller organic molecules

dehydration synthesis

(also, condensation) reaction that links monomer molecules together, releasing a molecule of water for each bond formed

hydrolysis

reaction causes breakdown of larger molecules into smaller molecules with the utilization of water

monomer

smallest unit of larger molecules called polymers

polymer

chain of monomer residues that is linked by covalent bonds;

polymerization is the process of polymer formation from monomers by condensation

Carbohydrates

By the end of this section, you will be able to:

- Discuss the role of carbohydrates in cells and in the extracellular materials of animals and plants
- Explain the classifications of carbohydrates
- List common monosaccharides, disaccharides, and polysaccharides

Most people are familiar with carbohydrates, one type of macromolecule, especially when it comes to what we eat. To lose weight, some individuals adhere to “low-carb” diets. Athletes, in contrast, often “carb-load” before important competitions to ensure that they have enough energy to compete at a high level. Carbohydrates are, in fact, an essential part of our diet; grains, fruits, and vegetables are all natural sources of carbohydrates. Carbohydrates provide energy to the body, particularly through glucose, a simple sugar that is a component of **starch** and an ingredient in many staple foods. Carbohydrates also have other important functions in humans, animals, and plants.

Molecular Structures

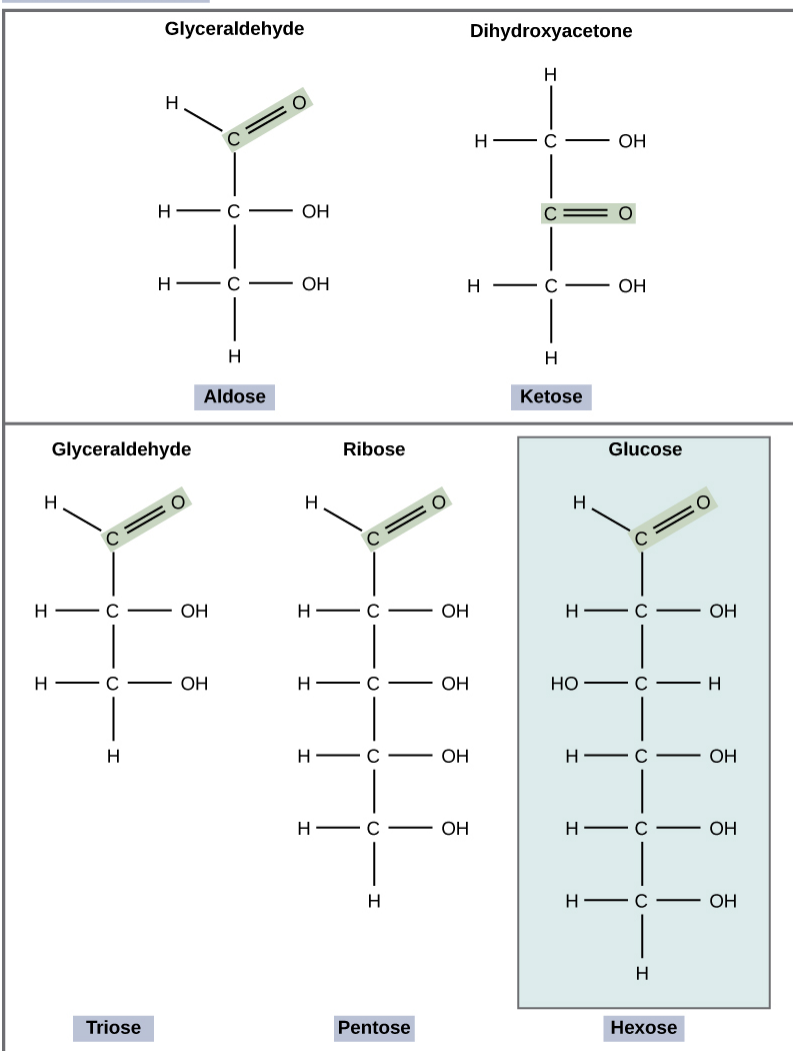
Carbohydrates can be represented by the stoichiometric formula $(\text{CH}_2\text{O})_n$, where n is the number of carbons in the molecule. In other words, the ratio of carbon to hydrogen to oxygen is 1:2:1 in carbohydrate molecules. This formula also explains the origin of the term “carbohydrate”: the components are carbon (“carbo”) and the components of water (hence, “hydrate”). Carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides.

Monosaccharides

Monosaccharides (mono- = “one”; sacchar- = “sweet”) are simple sugars, the most common of which is glucose. In monosaccharides, the number of carbons usually ranges from three to seven. Most monosaccharide names end with the suffix -ose. If the sugar has an aldehyde group (the functional group with the structure R-CHO), it is known as an aldose, and if it has a

ketone group (the functional group with the structure $\text{RC}(=\text{O})\text{R}'$), it is known as a ketose. Depending on the number of carbons in the sugar, they also may be known as trioses (three carbons), pentoses (five carbons), and or hexoses (six carbons). See [\[link\]](#) for an illustration of the monosaccharides.

MONOSACCHARIDES



Monosaccharides are classified based on the position of their carbonyl group and the number of carbons in the backbone. Aldoses have a carbonyl group (indicated in green) at the end of the carbon chain, and ketoses

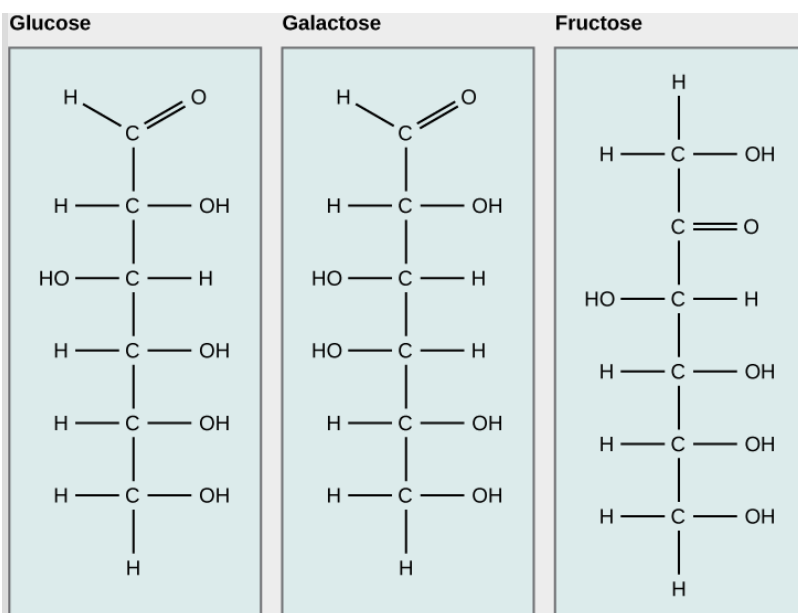
have a carbonyl group in the middle of the carbon chain. Trioses, pentoses, and hexoses have three, five, and six carbon backbones, respectively.

The chemical formula for glucose is $C_6H_{12}O_6$. In humans, glucose is an important source of energy. During cellular respiration, energy is released from glucose, and that energy is used to help make adenosine triphosphate (ATP). Plants synthesize glucose using carbon dioxide and water, and glucose in turn is used for energy requirements for the plant. Excess glucose is often stored as starch that is catabolized (the breakdown of larger molecules by cells) by humans and other animals that feed on plants.

Galactose (part of lactose, or milk sugar) and fructose (found in sucrose, in fruit) are other common monosaccharides. Although glucose, galactose, and fructose all have the same chemical formula ($C_6H_{12}O_6$), they differ structurally and chemically (and are known as isomers) because of the different arrangement of functional groups around the asymmetric carbon; all of these monosaccharides have more than one asymmetric carbon ([link](#)).

Note:

Art Connection

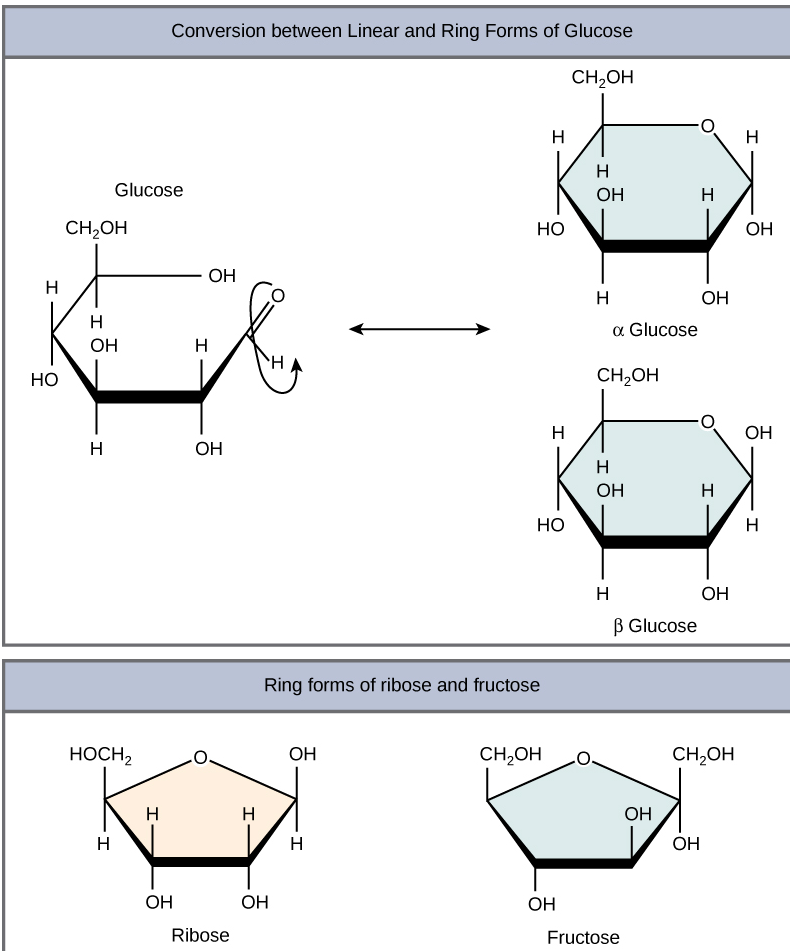


Glucose, galactose, and fructose are all hexoses. They are structural isomers, meaning they have the same chemical formula ($\text{C}_6\text{H}_{12}\text{O}_6$) but a different arrangement of atoms.

What kind of sugars are these, aldose or ketose?

Glucose, galactose, and fructose are isomeric monosaccharides (hexoses), meaning they have the same chemical formula but have slightly different structures. Glucose and galactose are aldoses, and fructose is a ketose.

Monosaccharides can exist as a linear chain or as ring-shaped molecules; in aqueous solutions they are usually found in ring forms ([link](#)). Glucose in a ring form can have two different arrangements of the hydroxyl group (OH) around the anomeric carbon (carbon 1 that becomes asymmetric in the process of ring formation). If the hydroxyl group is below carbon number 1 in the sugar, it is said to be in the alpha (α) position, and if it is above the plane, it is said to be in the beta (β) position.

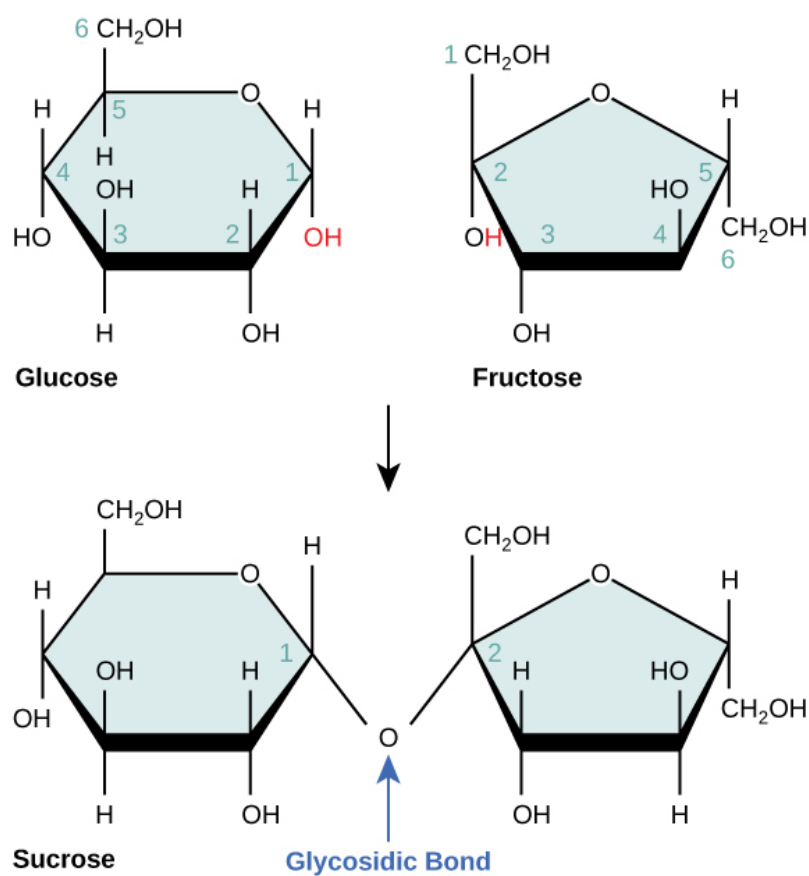


Five and six carbon monosaccharides exist in equilibrium between linear and ring forms. When the ring forms, the side chain it closes on is locked into an α or β position.

Fructose and ribose also form rings, although they form five-membered rings as opposed to the six-membered ring of glucose.

Disaccharides

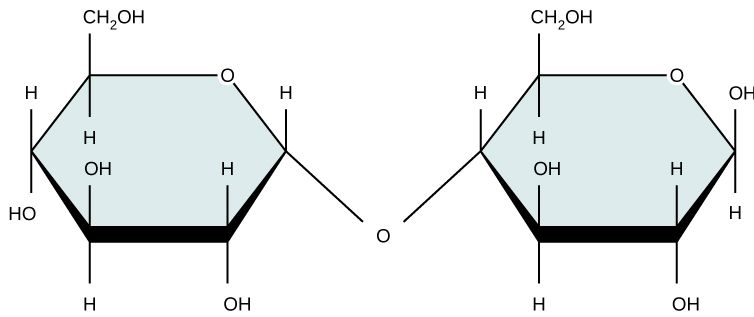
Disaccharides (di- = “two”) form when two monosaccharides undergo a dehydration reaction (also known as a condensation reaction or dehydration synthesis). During this process, the hydroxyl group of one monosaccharide combines with the hydrogen of another monosaccharide, releasing a molecule of water and forming a covalent bond. A covalent bond formed between a carbohydrate molecule and another molecule (in this case, between two monosaccharides) is known as a **glycosidic bond** ([\[link\]](#)). Glycosidic bonds (also called glycosidic linkages) can be of the alpha or the beta type.



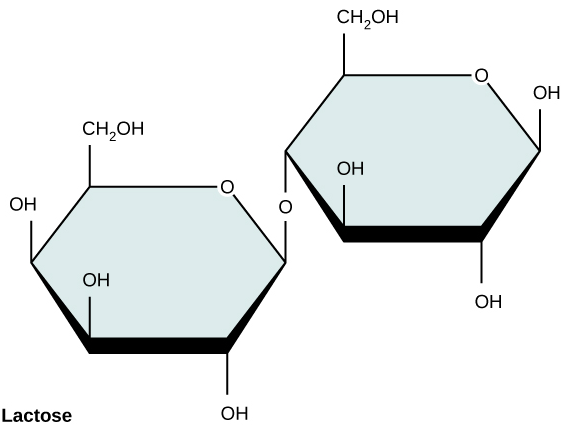
Sucrose is formed when a monomer of glucose and a monomer of fructose are joined in a dehydration reaction to form a glycosidic bond. In the process, a water molecule is lost. By convention, the carbon

atoms in a monosaccharide are numbered from the terminal carbon closest to the carbonyl group. In sucrose, a glycosidic linkage is formed between carbon 1 in glucose and carbon 2 in fructose.

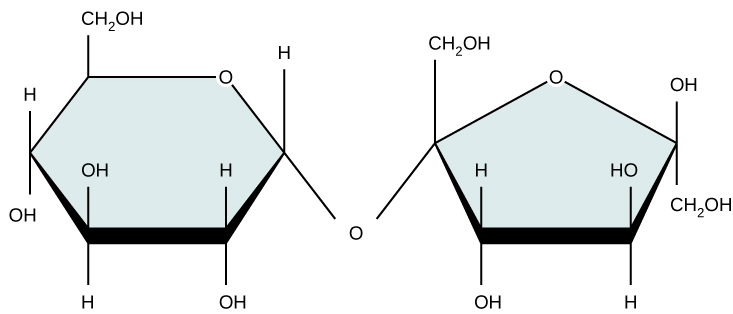
Common disaccharides include lactose, maltose, and sucrose ([\[link\]](#)). Lactose is a disaccharide consisting of the monomers glucose and galactose. It is found naturally in milk. Maltose, or malt sugar, is a disaccharide formed by a dehydration reaction between two glucose molecules. The most common disaccharide is sucrose, or table sugar, which is composed of the monomers glucose and fructose.



Maltose



Lactose



Sucrose

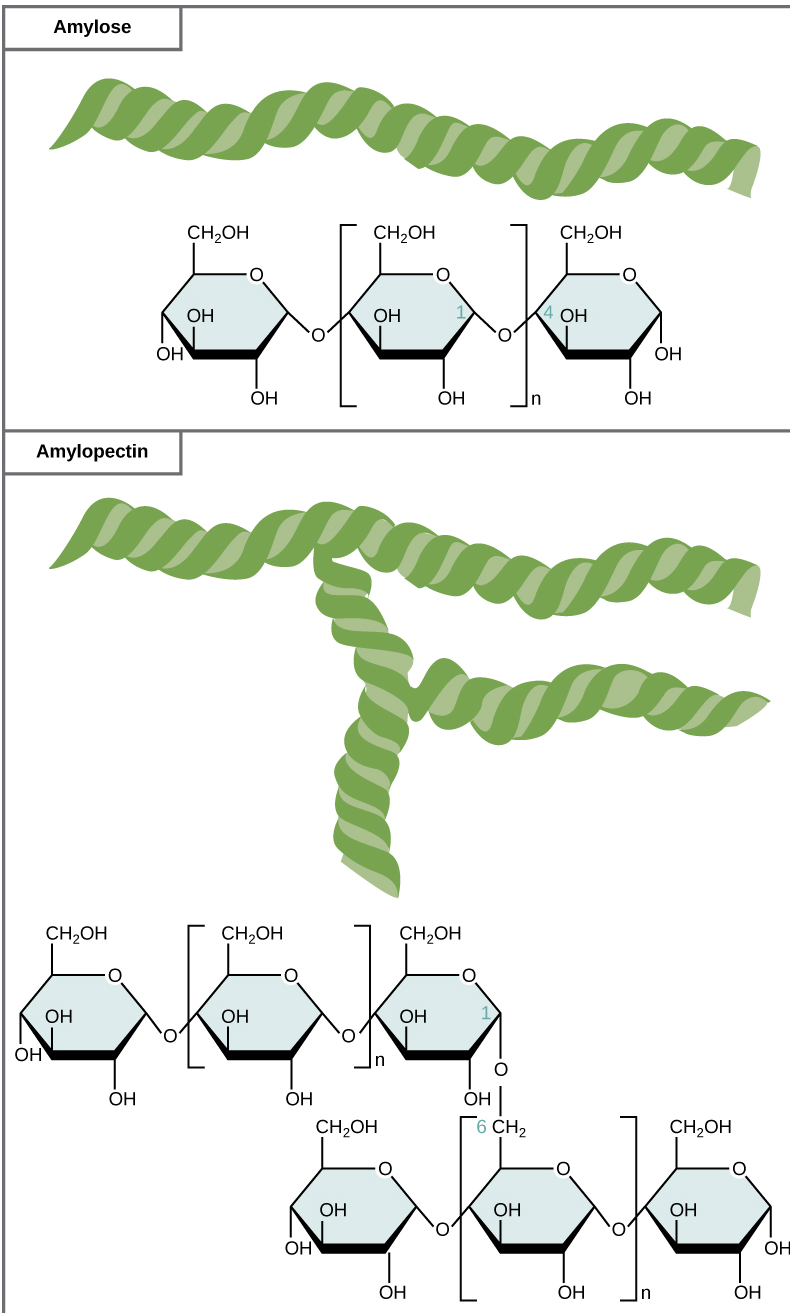
Common disaccharides include maltose (grain sugar), lactose (milk sugar), and sucrose (table sugar).

Polysaccharides

A long chain of monosaccharides linked by glycosidic bonds is known as a **polysaccharide** (poly- = “many”). The chain may be branched or unbranched, and it may contain different types of monosaccharides. The molecular weight may be 100,000 daltons or more depending on the number of monomers joined. Starch, glycogen, cellulose, and chitin are primary examples of polysaccharides.

Starch is the stored form of sugars in plants and is made up of a mixture of amylose and amylopectin (both polymers of glucose). Plants are able to synthesize glucose, and the excess glucose, beyond the plant’s immediate energy needs, is stored as starch in different plant parts, including roots and seeds. The starch in the seeds provides food for the embryo as it germinates and can also act as a source of food for humans and animals. The starch that is consumed by humans is broken down by enzymes, such as salivary amylases, into smaller molecules, such as maltose and glucose. The cells can then absorb the glucose.

Starch is made up of glucose monomers that are joined by α 1-4 or α 1-6 glycosidic bonds. The numbers 1-4 and 1-6 refer to the carbon number of the two residues that have joined to form the bond. As illustrated in [\[link\]](#), amylose is starch formed by unbranched chains of glucose monomers (only α 1-4 linkages), whereas amylopectin is a branched polysaccharide (α 1-6 linkages at the branch points).



Amylose and amylopectin are two different forms of starch. Amylose is composed of unbranched chains of glucose monomers connected by α 1,4 glycosidic linkages.

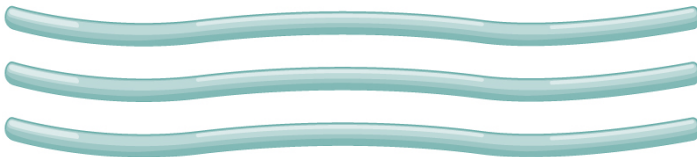
Amylopectin is composed of branched chains of glucose monomers connected by α 1,4 and α 1,6 glycosidic linkages. Because

of the way the subunits are joined, the glucose chains have a helical structure. Glycogen (not shown) is similar in structure to amylopectin but more highly branched.

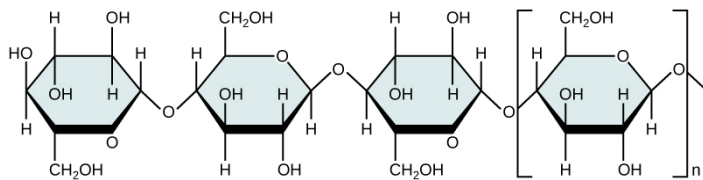
Glycogen is the storage form of glucose in humans and other vertebrates and is made up of monomers of glucose. Glycogen is the animal equivalent of starch and is a highly branched molecule usually stored in liver and muscle cells. Whenever blood glucose levels decrease, glycogen is broken down to release glucose in a process known as glycogenolysis.

Cellulose is the most abundant natural biopolymer. The cell wall of plants is mostly made of cellulose; this provides structural support to the cell. Wood and paper are mostly cellulosic in nature. Cellulose is made up of glucose monomers that are linked by β 1-4 glycosidic bonds ([\[link\]](#)).

Cellulose fibers



Cellulose structure



In cellulose, glucose monomers are linked in unbranched chains by β 1-4 glycosidic linkages. Because of the way the glucose subunits are joined, every glucose monomer is flipped relative to the next one resulting in a linear, fibrous structure.

As shown in [\[link\]](#), every other glucose monomer in cellulose is flipped over, and the monomers are packed tightly as extended long chains. This gives cellulose its rigidity and high tensile strength—which is so important to plant cells. While the β 1-4 linkage cannot be broken down by human digestive enzymes, herbivores such as cows, koalas, and buffalos are able, with the help of the specialized flora in their stomach, to digest plant material that is rich in cellulose and use it as a food source. In these animals, certain species of bacteria and protists reside in the rumen (part of the digestive system of herbivores) and secrete the enzyme cellulase. The appendix of grazing animals also contains bacteria that digest cellulose, giving it an important role in the digestive systems of ruminants. Cellulases can break down cellulose into glucose monomers that can be used as an energy source by the animal. Termites are also able to break down cellulose because of the presence of other organisms in their bodies that secrete cellulases.

Carbohydrates serve various functions in different animals. Arthropods (insects, crustaceans, and others) have an outer skeleton, called the exoskeleton, which protects their internal body parts (as seen in the bee in [\[link\]](#)). This exoskeleton is made of the biological macromolecule **chitin**, which is a polysaccharide-containing nitrogen. It is made of repeating units of N-acetyl- β -d-glucosamine, a modified sugar. Chitin is also a major component of fungal cell walls; fungi are neither animals nor plants and form a kingdom of their own in the domain Eukarya.



Insects have a hard outer exoskeleton made of chitin, a type of polysaccharide. (credit: Louise Docker)

Note:

Career Connections

Registered Dietitian

Obesity is a worldwide health concern, and many diseases such as diabetes and heart disease are becoming more prevalent because of obesity. This is one of the reasons why registered dietitians are increasingly sought after for advice. Registered dietitians help plan nutrition programs for individuals in various settings. They often work with patients in health care facilities, designing nutrition plans to treat and prevent diseases. For example, dietitians may teach a patient with diabetes how to manage blood sugar levels by eating the correct types and amounts of carbohydrates. Dietitians may also work in nursing homes, schools, and private practices. To become a registered dietitian, one needs to earn at least a bachelor's degree in dietetics, nutrition, food technology, or a related field. In addition, registered dietitians must complete a supervised internship

program and pass a national exam. Those who pursue careers in dietetics take courses in nutrition, chemistry, biochemistry, biology, microbiology, and human physiology. Dietitians must become experts in the chemistry and physiology (biological functions) of food (proteins, carbohydrates, and fats).

Benefits of Carbohydrates

Are carbohydrates good for you? People who wish to lose weight are often told that carbohydrates are bad for them and should be avoided. Some diets completely forbid carbohydrate consumption, claiming that a low-carbohydrate diet helps people to lose weight faster. However, carbohydrates have been an important part of the human diet for thousands of years; artifacts from ancient civilizations show the presence of wheat, rice, and corn in our ancestors' storage areas.

Carbohydrates should be supplemented with proteins, vitamins, and fats to be parts of a well-balanced diet. Calorie-wise, a gram of carbohydrate provides 4.3 Kcal. For comparison, fats provide 9 Kcal/g, a less desirable ratio. Carbohydrates contain soluble and insoluble elements; the insoluble part is known as fiber, which is mostly cellulose. Fiber has many uses; it promotes regular bowel movement by adding bulk, and it regulates the rate of consumption of blood glucose. Fiber also helps to remove excess cholesterol from the body: fiber binds to the cholesterol in the small intestine, then attaches to the cholesterol and prevents the cholesterol particles from entering the bloodstream, and then cholesterol exits the body via the feces. Fiber-rich diets also have a protective role in reducing the occurrence of colon cancer. In addition, a meal containing whole grains and vegetables gives a feeling of fullness. As an immediate source of energy, glucose is broken down during the process of cellular respiration, which produces ATP, the energy currency of the cell. Without the consumption of carbohydrates, the availability of "instant energy" would be reduced. Eliminating carbohydrates from the diet is not the best way to lose weight. A low-calorie diet that is rich in whole grains, fruits, vegetables, and lean

meat, together with plenty of exercise and plenty of water, is the more sensible way to lose weight.

Note:

Link to Learning



For an additional perspective on carbohydrates, explore “Biomolecules: the Carbohydrates” through this [interactive animation](#).

Section Summary

Carbohydrates are a group of macromolecules that are a vital energy source for the cell and provide structural support to plant cells, fungi, and all of the arthropods that include lobsters, crabs, shrimp, insects, and spiders. Carbohydrates are classified as monosaccharides, disaccharides, and polysaccharides depending on the number of monomers in the molecule. Monosaccharides are linked by glycosidic bonds that are formed as a result of dehydration reactions, forming disaccharides and polysaccharides with the elimination of a water molecule for each bond formed. Glucose, galactose, and fructose are common monosaccharides, whereas common disaccharides include lactose, maltose, and sucrose. Starch and glycogen, examples of polysaccharides, are the storage forms of glucose in plants and animals, respectively. The long polysaccharide chains may be branched or unbranched. Cellulose is an example of an unbranched polysaccharide, whereas amylopectin, a constituent of starch, is a highly branched molecule. Storage of glucose, in the form of polymers like starch or glycogen, makes it slightly less accessible for metabolism; however, this prevents it from

leaking out of the cell or creating a high osmotic pressure that could cause excessive water uptake by the cell.

Art Connections

Exercise:

Problem: [\[link\]](#) What kind of sugars are these, aldose or ketose?

Solution:

[\[link\]](#) Glucose and galactose are aldoses. Fructose is a ketose.

Review Questions

Exercise:

Problem: An example of a monosaccharide is _____.

- a. fructose
- b. glucose
- c. galactose
- d. all of the above

Solution:

D

Exercise:

Problem: Cellulose and starch are examples of:

- a. monosaccharides
- b. disaccharides
- c. lipids

d. polysaccharides

Solution:

D

Exercise:

Problem:

Plant cell walls contain which of the following in abundance?

- a. starch
 - b. cellulose
 - c. glycogen
 - d. lactose
-

Solution:

B

Exercise:

Problem:

Lactose is a disaccharide formed by the formation of a _____ bond between glucose and _____.

- a. glycosidic; lactose
 - b. glycosidic; galactose
 - c. hydrogen; sucrose
 - d. hydrogen; fructose
-

Solution:

B

Free Response

Exercise:

Problem:

Describe the similarities and differences between glycogen and starch.

Solution:

Glycogen and starch are polysaccharides. They are the storage form of glucose. Glycogen is stored in animals in the liver and in muscle cells, whereas starch is stored in the roots, seeds, and leaves of plants. Starch has two different forms, one unbranched (amylose) and one branched (amylopectin), whereas glycogen is a single type of a highly branched molecule.

Exercise:

Problem:

Why is it impossible for humans to digest food that contains cellulose?

Solution:

The β 1-4 glycosidic linkage in cellulose cannot be broken down by human digestive enzymes. Herbivores such as cows, koalas, and buffalos are able to digest grass that is rich in cellulose and use it as a food source because bacteria and protists in their digestive systems, especially in the rumen, secrete the enzyme cellulase. Cellulases can break down cellulose into glucose monomers that can be used as an energy source by the animal.

Glossary

carbohydrate

biological macromolecule in which the ratio of carbon to hydrogen and to oxygen is 1:2:1; carbohydrates serve as energy sources and

structural support in cells and form the a cellular exoskeleton of arthropods

cellulose

polysaccharide that makes up the cell wall of plants; provides structural support to the cell

chitin

type of carbohydrate that forms the outer skeleton of all arthropods that include crustaceans and insects; it also forms the cell walls of fungi

disaccharide

two sugar monomers that are linked together by a glycosidic bond

glycogen

storage carbohydrate in animals

glycosidic bond

bond formed by a dehydration reaction between two monosaccharides with the elimination of a water molecule

monosaccharide

single unit or monomer of carbohydrates

polysaccharide

long chain of monosaccharides; may be branched or unbranched

starch

storage carbohydrate in plants

Lipids

By the end of this section, you will be able to:

- Describe the four major types of lipids
- Explain the role of fats in storing energy
- Differentiate between saturated and unsaturated fatty acids
- Describe phospholipids and their role in cells
- Define the basic structure of a steroid and some functions of steroids
- Explain the how cholesterol helps to maintain the fluid nature of the plasma membrane

Lipids include a diverse group of compounds that are largely nonpolar in nature. This is because they are hydrocarbons that include mostly nonpolar carbon–carbon or carbon–hydrogen bonds. Non-polar molecules are hydrophobic (“water fearing”), or insoluble in water. Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of fats. Lipids also provide insulation from the environment for plants and animals ([\[link\]](#)). For example, they help keep aquatic birds and mammals dry when forming a protective layer over fur or feathers because of their water-repellant hydrophobic nature. Lipids are also the building blocks of many hormones and are an important constituent of all cellular membranes. Lipids include fats, oils, waxes, phospholipids, and steroids.



Hydrophobic lipids in the fur
of aquatic mammals, such as

this river otter, protect them
from the elements. (credit:
Ken Bosma)

Fats and Oils

A fat molecule consists of two main components—glycerol and fatty acids. Glycerol is an organic compound (alcohol) with three carbons, five hydrogens, and three hydroxyl (OH) groups. Fatty acids have a long chain of hydrocarbons to which a carboxyl group is attached, hence the name “fatty acid.” The number of carbons in the fatty acid may range from 4 to 36; most common are those containing 12–18 carbons. In a fat molecule, the fatty acids are attached to each of the three carbons of the glycerol molecule with an ester bond through an oxygen atom ([\[link\]](#)).

$$\begin{array}{c} \text{H} \\ | \\ \text{H}-\text{C}-\text{OH} \\ | \\ \text{H}-\text{C}-\text{OH} \\ | \\ \text{H}-\text{C}-\text{OH} \\ | \\ \text{H} \end{array}$$
CCCCCCCCCCCCCCCCO

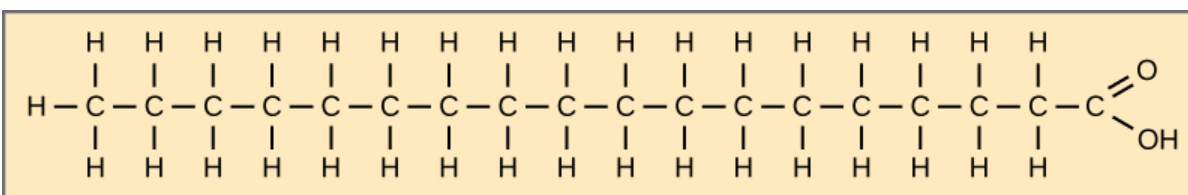
Diagram illustrating the chemical structures of three saturated fatty acids, showing the carboxyl group (left) and the hydrocarbon chain (right).

- Top structure (Stearic acid):** A long, straight hydrocarbon chain with 18 carbon atoms in total (including the carboxyl carbon).
- Middle structure (Myristic acid):** A medium-length, straight hydrocarbon chain with 14 carbon atoms in total (including the carboxyl carbon).
- Bottom structure (Lauric acid):** A shorter, straight hydrocarbon chain with 12 carbon atoms in total (including the carboxyl carbon).

During this ester bond formation, three water molecules are released. The three fatty acids in the triacylglycerol may be similar or dissimilar. Fats are also called **triacylglycerols** or **triglycerides** because of their chemical

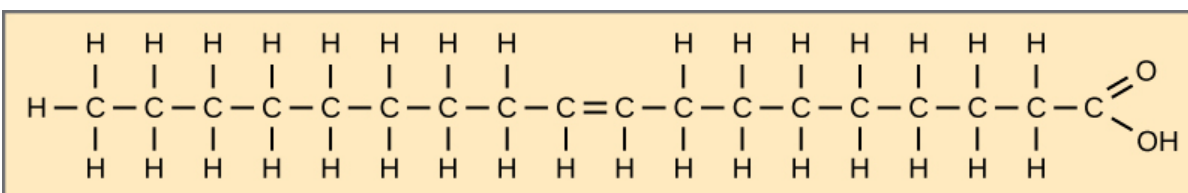
structure. Some fatty acids have common names that specify their origin. For example, palmitic acid, a **saturated fatty acid**, is derived from the palm tree. Arachidic acid is derived from *Arachis hypogea*, the scientific name for groundnuts or peanuts.

Fatty acids may be saturated or unsaturated. In a fatty acid chain, if there are only single bonds between neighboring carbons in the hydrocarbon chain, the fatty acid is said to be saturated. Saturated fatty acids are saturated with hydrogen; in other words, the number of hydrogen atoms attached to the carbon skeleton is maximized. Stearic acid is an example of a saturated fatty acid ([\[link\]](#))



Stearic acid is a common saturated fatty acid.

When the hydrocarbon chain contains a double bond, the fatty acid is said to be **unsaturated**. Oleic acid is an example of an unsaturated fatty acid ([\[link\]](#)).



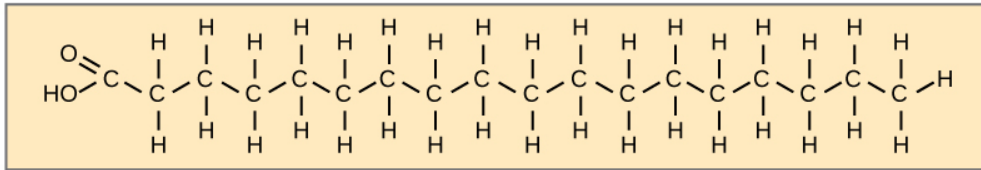
Oleic acid is a common unsaturated fatty acid.

Most unsaturated fats are liquid at room temperature and are called oils. If there is one double bond in the molecule, then it is known as a monounsaturated fat (e.g., olive oil), and if there is more than one double bond, then it is known as a polyunsaturated fat (e.g., canola oil).

When a fatty acid has no double bonds, it is known as a saturated fatty acid because no more hydrogen may be added to the carbon atoms of the chain. A fat may contain similar or different fatty acids attached to glycerol. Long straight fatty acids with single bonds tend to get packed tightly and are solid at room temperature. Animal fats with stearic acid and palmitic acid (common in meat) and the fat with butyric acid (common in butter) are examples of saturated fats. Mammals store fats in specialized cells called adipocytes, where globules of fat occupy most of the cell's volume. In plants, fat or oil is stored in many seeds and is used as a source of energy during seedling development. Unsaturated fats or oils are usually of plant origin and contain *cis* unsaturated fatty acids. *Cis* and *trans* indicate the configuration of the molecule around the double bond. If hydrogens are present in the same plane, it is referred to as a *cis* fat; if the hydrogen atoms are on two different planes, it is referred to as a **trans fat**. The *cis* double bond causes a bend or a “kink” that prevents the fatty acids from packing tightly, keeping them liquid at room temperature ([\[link\]](#)). Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats. Unsaturated fats help to lower blood cholesterol levels whereas saturated fats contribute to plaque formation in the arteries.

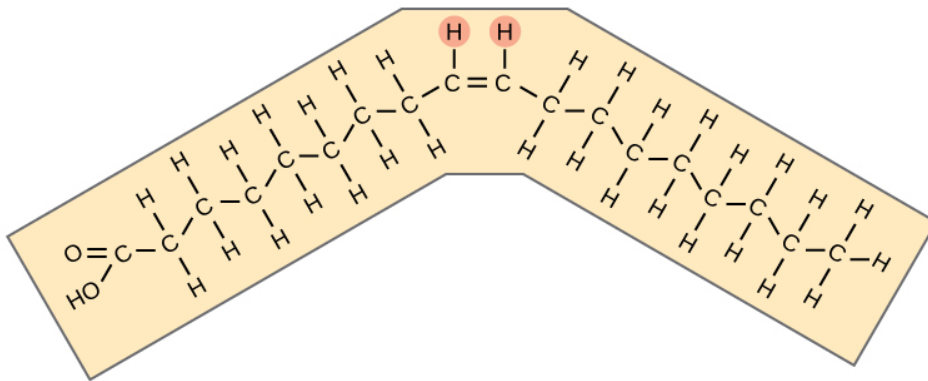
Saturated fatty acid

Stearic acid

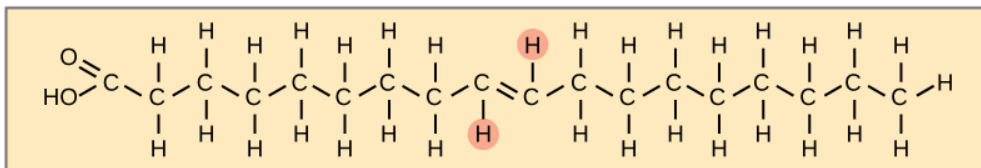


Unsaturated fatty acids

Cis oleic acid



Trans oleic acid



Saturated fatty acids have hydrocarbon chains connected by single bonds only. Unsaturated fatty acids have one or more double bonds. Each double bond may be in a *cis* or *trans* configuration. In the *cis* configuration, both hydrogens are on the same side of the hydrocarbon chain. In the *trans* configuration, the hydrogens are on opposite sides. A *cis* double bond causes a kink in the chain.

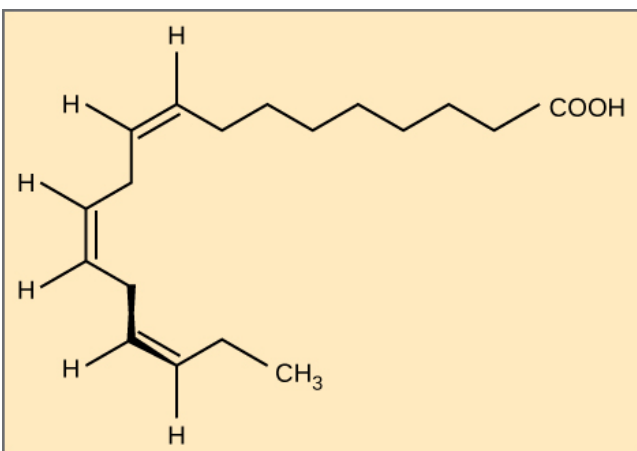
Trans Fats

In the food industry, oils are artificially hydrogenated to make them semi-solid and of a consistency desirable for many processed food products. Simply speaking, hydrogen gas is bubbled through oils to solidify them. During this hydrogenation process, double bonds of the *cis*- conformation in the hydrocarbon chain may be converted to double bonds in the *trans*-conformation.

Margarine, some types of peanut butter, and shortening are examples of artificially hydrogenated *trans* fats. Recent studies have shown that an increase in *trans* fats in the human diet may lead to an increase in levels of low-density lipoproteins (LDL), or “bad” cholesterol, which in turn may lead to plaque deposition in the arteries, resulting in heart disease. Many fast food restaurants have recently banned the use of *trans* fats, and food labels are required to display the *trans* fat content.

Omega Fatty Acids

Essential fatty acids are fatty acids required but not synthesized by the human body. Consequently, they have to be supplemented through ingestion via the diet. **Omega-3** fatty acids (like that shown in [\[link\]](#)) fall into this category and are one of only two known for humans (the other being omega-6 fatty acid). These are polyunsaturated fatty acids and are called omega-3 because the third carbon from the end of the hydrocarbon chain is connected to its neighboring carbon by a double bond.



Alpha-linolenic acid is an example of an omega-3 fatty acid. It has three *cis* double bonds and, as a result, a curved shape. For clarity, the carbons are not shown. Each singly bonded carbon has two hydrogens associated with it, also not shown.

The farthest carbon away from the carboxyl group is numbered as the omega (ω) carbon, and if the double bond is between the third and fourth carbon from that end, it is known as an omega-3 fatty acid. Nutritionally important because the body does not make them, omega-3 fatty acids include alpha-linolenic acid (ALA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), all of which are polyunsaturated. Salmon, trout, and tuna are good sources of omega-3 fatty acids. Research indicates that omega-3 fatty acids reduce the risk of sudden death from heart attacks, reduce triglycerides in the blood, lower blood pressure, and prevent thrombosis by inhibiting blood clotting. They also reduce inflammation, and may help reduce the risk of some cancers in animals.

Like carbohydrates, fats have received a lot of bad publicity. It is true that eating an excess of fried foods and other “fatty” foods leads to weight gain. However, fats do have important functions. Many vitamins are fat soluble,

and fats serve as a long-term storage form of fatty acids: a source of energy. They also provide insulation for the body. Therefore, “healthy” fats in moderate amounts should be consumed on a regular basis.

Waxes

Wax covers the feathers of some aquatic birds and the leaf surfaces of some plants. Because of the hydrophobic nature of waxes, they prevent water from sticking on the surface ([\[link\]](#)). Waxes are made up of long fatty acid chains esterified to long-chain alcohols.

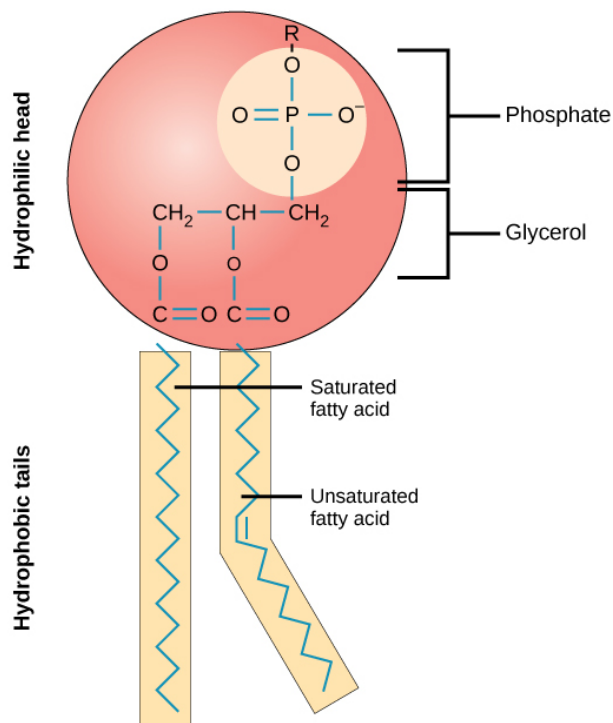


Waxy coverings on some leaves are made of lipids. (credit: Roger Griffith)

Phospholipids

Phospholipids are major constituents of the plasma membrane, the outermost layer of animal cells. Like fats, they are composed of fatty acid chains attached to a glycerol or sphingosine backbone. Instead of three fatty acids attached as in triglycerides, however, there are two fatty acids forming

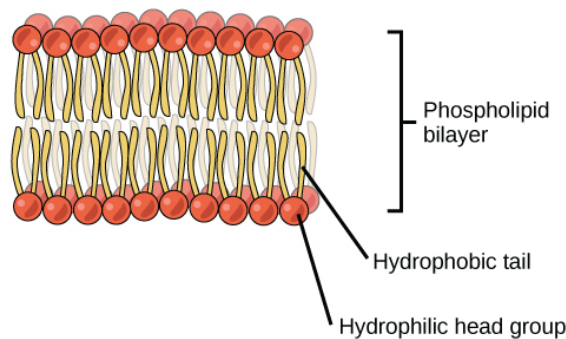
diacylglycerol, and the third carbon of the glycerol backbone is occupied by a modified phosphate group ([\[link\]](#)). A phosphate group alone attached to a diacylglycerol does not qualify as a phospholipid; it is phosphatidate (diacylglycerol 3-phosphate), the precursor of phospholipids. The phosphate group is modified by an alcohol. Phosphatidylcholine and phosphatidylserine are two important phospholipids that are found in plasma membranes.



A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. The phosphate may be modified by the addition of charged or polar chemical groups. Two chemical groups that may modify the phosphate, choline and serine, are shown here. Both choline and serine attach to the phosphate group at

the position labeled R via the hydroxyl group indicated in green.

A phospholipid is an amphipathic molecule, meaning it has a hydrophobic and a hydrophilic part. The fatty acid chains are hydrophobic and cannot interact with water, whereas the phosphate-containing group is hydrophilic and interacts with water ([\[link\]](#)).



The phospholipid bilayer is the major component of all cellular membranes. The hydrophilic head groups of the phospholipids face the aqueous solution. The hydrophobic tails are sequestered in the middle of the bilayer.

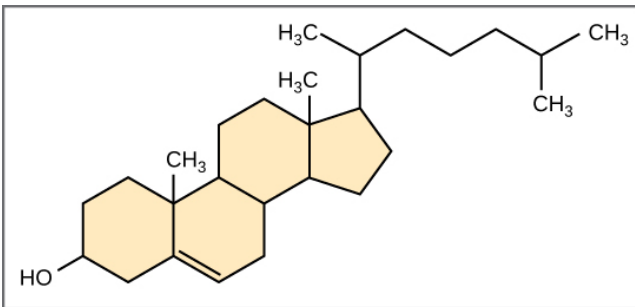
The head is the hydrophilic part, and the tail contains the hydrophobic fatty acids. In a membrane, a bilayer of phospholipids forms the matrix of the structure, the fatty acid tails of phospholipids face inside, away from water, whereas the phosphate group faces the outside, aqueous side ([\[link\]](#)).

Phospholipids are responsible for the dynamic nature of the plasma membrane. If a drop of phospholipids is placed in water, it spontaneously

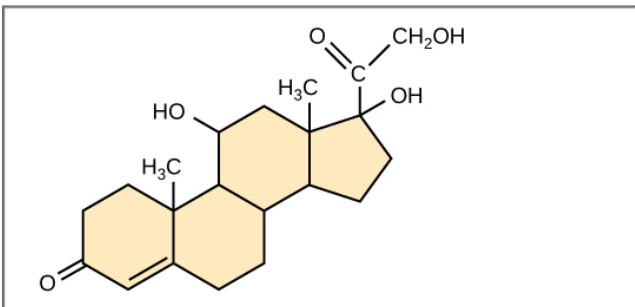
forms a structure known as a micelle, where the hydrophilic phosphate heads face the outside and the fatty acids face the interior of this structure.

Steroids

Unlike the phospholipids and fats discussed earlier, **steroids** have a fused ring structure. Although they do not resemble the other lipids, they are grouped with them because they are also hydrophobic and insoluble in water. All steroids have four linked carbon rings and several of them, like cholesterol, have a short tail ([link](#)). Many steroids also have the -OH functional group, which puts them in the alcohol classification (sterols).



Cholesterol



Cortisol

Steroids such as cholesterol and cortisol are composed of four fused hydrocarbon rings.

Cholesterol is the most common steroid. Cholesterol is mainly synthesized in the liver and is the precursor to many steroid hormones such as testosterone and estradiol, which are secreted by the gonads and endocrine glands. It is also the precursor to Vitamin D. Cholesterol is also the precursor of bile salts, which help in the emulsification of fats and their subsequent absorption by cells. Although cholesterol is often spoken of in negative terms by lay people, it is necessary for proper functioning of the body. It is a component of the plasma membrane of animal cells and is found within the phospholipid bilayer. Being the outermost structure in animal cells, the plasma membrane is responsible for the transport of materials and cellular recognition and it is involved in cell-to-cell communication.

Note:

Link to Learning



For an additional perspective on lipids, explore the interactive animation [“Biomolecules: The Lipids”](#).

Section Summary

Lipids are a class of macromolecules that are nonpolar and hydrophobic in nature. Major types include fats and oils, waxes, phospholipids, and steroids. Fats are a stored form of energy and are also known as triacylglycerols or triglycerides. Fats are made up of fatty acids and either glycerol or sphingosine. Fatty acids may be unsaturated or saturated,

depending on the presence or absence of double bonds in the hydrocarbon chain. If only single bonds are present, they are known as saturated fatty acids. Unsaturated fatty acids may have one or more double bonds in the hydrocarbon chain. Phospholipids make up the matrix of membranes. They have a glycerol or sphingosine backbone to which two fatty acid chains and a phosphate-containing group are attached. Steroids are another class of lipids. Their basic structure has four fused carbon rings. Cholesterol is a type of steroid and is an important constituent of the plasma membrane, where it helps to maintain the fluid nature of the membrane. It is also the precursor of steroid hormones such as testosterone.

Review Questions

Exercise:

Problem:

Saturated fats have all of the following characteristics except:

- a. they are solid at room temperature
- b. they have single bonds within the carbon chain
- c. they are usually obtained from animal sources
- d. they tend to dissolve in water easily

Solution:

D

Exercise:

Problem: Phospholipids are important components of _____.

- a. the plasma membrane of animal cells
 - b. the ring structure of steroids
 - c. the waxy covering on leaves
 - d. the double bond in hydrocarbon chains
-

Solution:

A

Free Response**Exercise:****Problem:**

Explain at least three functions that lipids serve in plants and/or animals.

Solution:

Fat serves as a valuable way for animals to store energy. It can also provide insulation. Waxes can protect plant leaves and mammalian fur from getting wet. Phospholipids and steroids are important components of animal cell membranes, as well as plant, fungal, and bacterial membranes.

Exercise:**Problem:**

Why have trans fats been banned from some restaurants? How are they created?

Solution:

Trans fats are created artificially when hydrogen gas is bubbled through oils to solidify them. The double bonds of the *cis* conformation in the hydrocarbon chain may be converted to double bonds in the *trans* configuration. Some restaurants are banning trans fats because they cause higher levels of LDL, or “bad” cholesterol.

Glossary

lipid

macromolecule that is nonpolar and insoluble in water

omega fat

type of polyunsaturated fat that is required by the body; the numbering of the carbon omega starts from the methyl end or the end that is farthest from the carboxylic end

phospholipid

major constituent of the membranes; composed of two fatty acids and a phosphate-containing group attached to a glycerol backbone

saturated fatty acid

long-chain of hydrocarbon with single covalent bonds in the carbon chain; the number of hydrogen atoms attached to the carbon skeleton is maximized

steroid

type of lipid composed of four fused hydrocarbon rings forming a planar structure

trans fat

fat formed artificially by hydrogenating oils, leading to a different arrangement of double bond(s) than those found in naturally occurring lipids

triacylglycerol (also, triglyceride)

fat molecule; consists of three fatty acids linked to a glycerol molecule

unsaturated fatty acid

long-chain hydrocarbon that has one or more double bonds in the hydrocarbon chain

wax

lipid made of a long-chain fatty acid that is esterified to a long-chain alcohol; serves as a protective coating on some feathers, aquatic mammal fur, and leaves

Proteins

By the end of this section, you will be able to:

- Describe the functions proteins perform in the cell and in tissues
- Discuss the relationship between amino acids and proteins
- Explain the four levels of protein organization
- Describe the ways in which protein shape and function are linked

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules.

Proteins may be structural, regulatory, contractile, or protective; they may serve in transport, storage, or membranes; or they may be toxins or enzymes. Each cell in a living system may contain thousands of proteins, each with a unique function. Their structures, like their functions, vary greatly. They are all, however, polymers of amino acids, arranged in a linear sequence.

Types and Functions of Proteins

Enzymes, which are produced by living cells, are catalysts in biochemical reactions (like digestion) and are usually complex or conjugated proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) it acts on. The enzyme may help in breakdown, rearrangement, or synthesis reactions. Enzymes that break down their substrates are called catabolic enzymes, enzymes that build more complex molecules from their substrates are called anabolic enzymes, and enzymes that affect the rate of reaction are called catalytic enzymes. It should be noted that all enzymes increase the rate of reaction and, therefore, are considered to be organic catalysts. An example of an enzyme is salivary amylase, which hydrolyzes its substrate amylose, a component of starch.

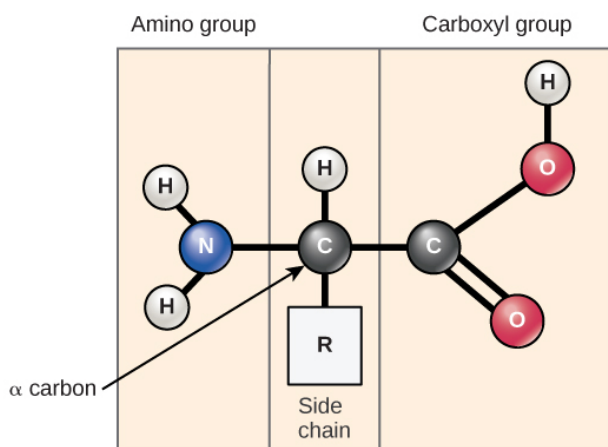
Hormones are chemical-signaling molecules, usually small proteins or steroids, secreted by endocrine cells that act to control or regulate specific physiological processes, including growth, development, metabolism, and reproduction. For example, insulin is a protein hormone that helps to regulate the blood glucose level. The primary types and functions of proteins are listed in [\[link\]](#).

Protein Types and Functions		
Type	Examples	Functions
Digestive Enzymes	Amylase, lipase, pepsin, trypsin	Help in digestion of food by catabolizing nutrients into monomeric units
Transport	Hemoglobin, albumin	Carry substances in the blood or lymph throughout the body
Structural	Actin, tubulin, keratin	Construct different structures, like the cytoskeleton
Hormones	Insulin, thyroxine	Coordinate the activity of different body systems
Defense	Immunoglobulins	Protect the body from foreign pathogens
Contractile	Actin, myosin	Effect muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide nourishment in early development of the embryo and the seedling

Proteins have different shapes and molecular weights; some proteins are globular in shape whereas others are fibrous in nature. For example, hemoglobin is a globular protein, but collagen, found in our skin, is a fibrous protein. Protein shape is critical to its function, and this shape is maintained by many different types of chemical bonds. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to loss of function, known as **denaturation**. All proteins are made up of different arrangements of the same 20 types of amino acids.

Amino Acids

Amino acids are the monomers that make up proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group (COOH), and to a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the central atom known as the R group ([\[link\]](#)).



Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached.

The name "amino acid" is derived from the fact that they contain both amino group and carboxyl-acid-group in their basic structure. As mentioned, there are 20 amino acids present in proteins. Nine of these are considered essential amino acids in humans because the human body cannot produce them and they are obtained from the diet. For each amino acid, the R group (or side chain) is different ([\[link\]](#)).

Note:**Art Connection**

AMINO ACID			
Nonpolar, aliphatic R groups			
	Glycine	Alanine	Valine
Polar, uncharged R groups			
	Serine	Threonine	Cysteine
	Proline	Asparagine	Glutamine

AMINO ACID			
Positively charged R groups			
	Lysine	Arginine	Histidine
Negatively charged R groups			
	Aspartate	Glutamate	
Nonpolar, aromatic R groups			
	Phenylalanine	Tyrosine	Tryptophan

There are 20 common amino acids commonly found in proteins, each with a different R group (variant group) that determines its chemical nature.

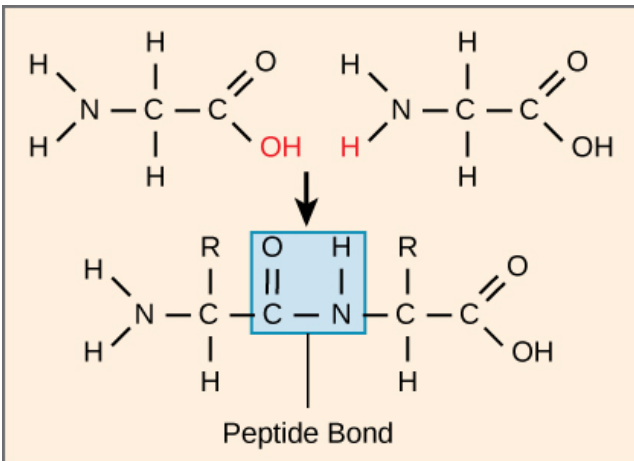
Which categories of amino acid would you expect to find on the surface of a soluble protein, and which would you expect to find in the interior? What distribution of amino acids would you expect to find in a protein embedded in a lipid bilayer?

The chemical nature of the side chain determines the nature of the amino acid (that is, whether it is acidic, basic, polar, or nonpolar). For example, the

amino acid glycine has a hydrogen atom as the R group. Amino acids such as valine, methionine, and alanine are nonpolar or hydrophobic in nature, while amino acids such as serine, threonine, and cysteine are polar and have hydrophilic side chains. The side chains of lysine and arginine are positively charged, and therefore these amino acids are also known as basic amino acids. Proline has an R group that is linked to the amino group, forming a ring-like structure. Proline is an exception to the standard structure of an amino acid since its amino group is not separate from the side chain ([\[link\]](#)).

Amino acids are represented by a single upper case letter or a three-letter abbreviation. For example, valine is known by the letter V or the three-letter symbol val. Just as some fatty acids are essential to a diet, some amino acids are necessary as well. They are known as essential amino acids, and in humans they include isoleucine, leucine, and cysteine. Essential amino acids refer to those necessary for construction of proteins in the body, although not produced by the body; which amino acids are essential varies from organism to organism.

The sequence and the number of amino acids ultimately determine the protein's shape, size, and function. Each amino acid is attached to another amino acid by a covalent bond, known as a **peptide bond**, which is formed by a dehydration reaction. The carboxyl group of one amino acid and the amino group of the incoming amino acid combine, releasing a molecule of water. The resulting bond is the peptide bond ([\[link\]](#)).



Peptide bond formation is a dehydration synthesis reaction. The carboxyl group of one amino acid is linked to the amino group of the incoming amino acid. In the process, a molecule of water is released.

The products formed by such linkages are called peptides. As more amino acids join to this growing chain, the resulting chain is known as a polypeptide. Each polypeptide has a free amino group at one end. This end is called the N terminal, or the amino terminal, and the other end has a free carboxyl group, also known as the C or carboxyl terminal. While the terms polypeptide and protein are sometimes used interchangeably, a polypeptide is technically a polymer of amino acids, whereas the term protein is used for a polypeptide or polypeptides that have combined together, often have bound non-peptide prosthetic groups, have a distinct shape, and have a unique function. After protein synthesis (translation), most proteins are modified. These are known as post-translational modifications. They may undergo cleavage, phosphorylation, or may require the addition of other chemical groups. Only after these modifications is the protein completely functional.

Note:

Link to Learning



Click through the steps of protein synthesis in this [interactive tutorial](#).

Note:

Evolution Connection

The Evolutionary Significance of Cytochrome c

Cytochrome c is an important component of the electron transport chain, a part of cellular respiration, and it is normally found in the cellular organelle, the mitochondrion. This protein has a heme prosthetic group, and the central ion of the heme gets alternately reduced and oxidized during electron transfer. Because this essential protein's role in producing cellular energy is crucial, it has changed very little over millions of years. Protein sequencing has shown that there is a considerable amount of cytochrome c amino acid sequence homology among different species; in other words, evolutionary kinship can be assessed by measuring the similarities or differences among various species' DNA or protein sequences.

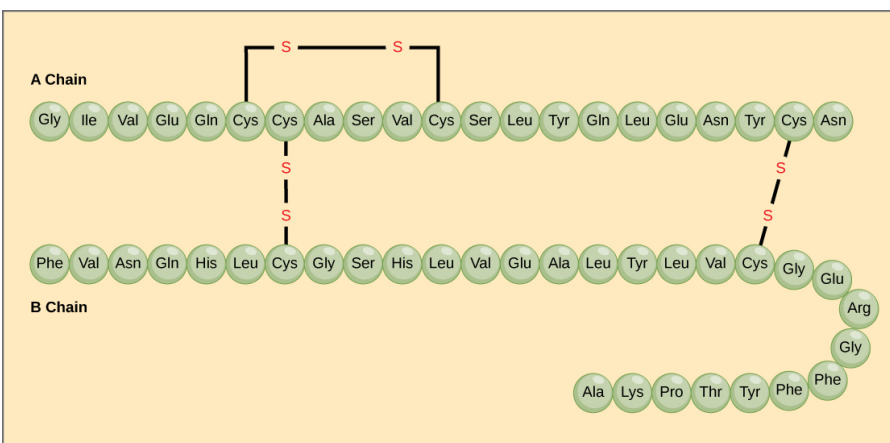
Scientists have determined that human cytochrome c contains 104 amino acids. For each cytochrome c molecule from different organisms that has been sequenced to date, 37 of these amino acids appear in the same position in all samples of cytochrome c. This indicates that there may have been a common ancestor. On comparing the human and chimpanzee protein sequences, no sequence difference was found. When human and rhesus monkey sequences were compared, the single difference found was in one amino acid. In another comparison, human to yeast sequencing shows a difference in the 44th position.

Protein Structure

As discussed earlier, the shape of a protein is critical to its function. For example, an enzyme can bind to a specific substrate at a site known as the active site. If this active site is altered because of local changes or changes in overall protein structure, the enzyme may be unable to bind to the substrate. To understand how the protein gets its final shape or conformation, we need to understand the four levels of protein structure: primary, secondary, tertiary, and quaternary.

Primary Structure

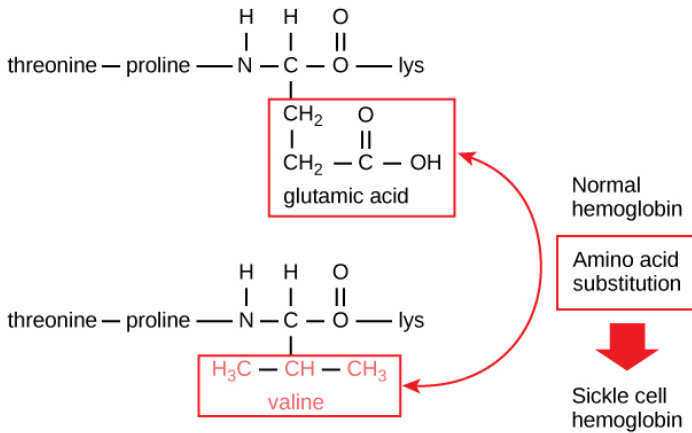
The unique sequence of amino acids in a polypeptide chain is its **primary structure**. For example, the pancreatic hormone insulin has two polypeptide chains, A and B, and they are linked together by disulfide bonds. The N terminal amino acid of the A chain is glycine, whereas the C terminal amino acid is asparagine ([\[link\]](#)). The sequences of amino acids in the A and B chains are unique to insulin.



Bovine serum insulin is a protein hormone made of two peptide chains, A (21 amino acids long) and B (30 amino acids long). In each chain, primary structure is indicated by three-letter abbreviations

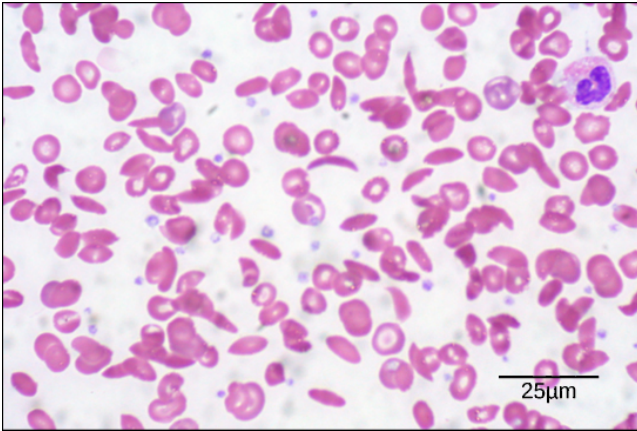
that represent the names of the amino acids in the order they are present. The amino acid cysteine (cys) has a sulfhydryl (SH) group as a side chain. Two sulfhydryl groups can react in the presence of oxygen to form a disulfide (S-S) bond. Two disulfide bonds connect the A and B chains together, and a third helps the A chain fold into the correct shape. Note that all disulfide bonds are the same length, but are drawn different sizes for clarity.

The unique sequence for every protein is ultimately determined by the gene encoding the protein. A change in nucleotide sequence of the gene's coding region may lead to a different amino acid being added to the growing polypeptide chain, causing a change in protein structure and function. In sickle cell anemia, the hemoglobin β chain (a small portion of which is shown in [\[link\]](#)) has a single amino acid substitution, causing a change in protein structure and function. Specifically, the amino acid glutamic acid is substituted by valine in the β chain. What is most remarkable to consider is that a hemoglobin molecule is made up of two alpha chains and two beta chains that each consist of about 150 amino acids. The molecule, therefore, has about 600 amino acids. The structural difference between a normal hemoglobin molecule and a sickle cell molecule—which dramatically decreases life expectancy—is a single amino acid of the 600. What is even more remarkable is that those 600 amino acids are encoded by three nucleotides each, and the mutation is caused by a single base change (point mutation), 1 in 1800 bases.



The beta chain of hemoglobin is 147 residues in length, yet a single amino acid substitution leads to sickle cell anemia. In normal hemoglobin, the amino acid at position seven is glutamate. In sickle cell hemoglobin, this glutamate is replaced by a valine.

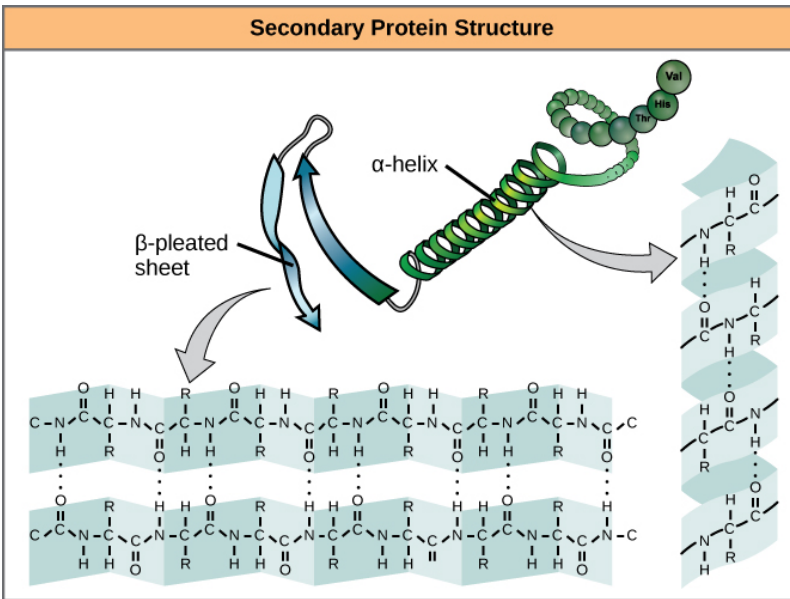
Because of this change of one amino acid in the chain, hemoglobin molecules form long fibers that distort the biconcave, or disc-shaped, red blood cells and assume a crescent or “sickle” shape, which clogs arteries ([link](#)). This can lead to myriad serious health problems such as breathlessness, dizziness, headaches, and abdominal pain for those affected by this disease.



In this blood smear, visualized at 535x magnification using bright field microscopy, sickle cells are crescent shaped, while normal cells are disc-shaped. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

Secondary Structure

The local folding of the polypeptide in some regions gives rise to the **secondary structure** of the protein. The most common are the **α -helix** and **β -pleated sheet** structures ([\[link\]](#)). Both structures are the α -helix structure—the helix held in shape by hydrogen bonds. The hydrogen bonds form between the oxygen atom in the carbonyl group in one amino acid and another amino acid that is four amino acids farther along the chain.

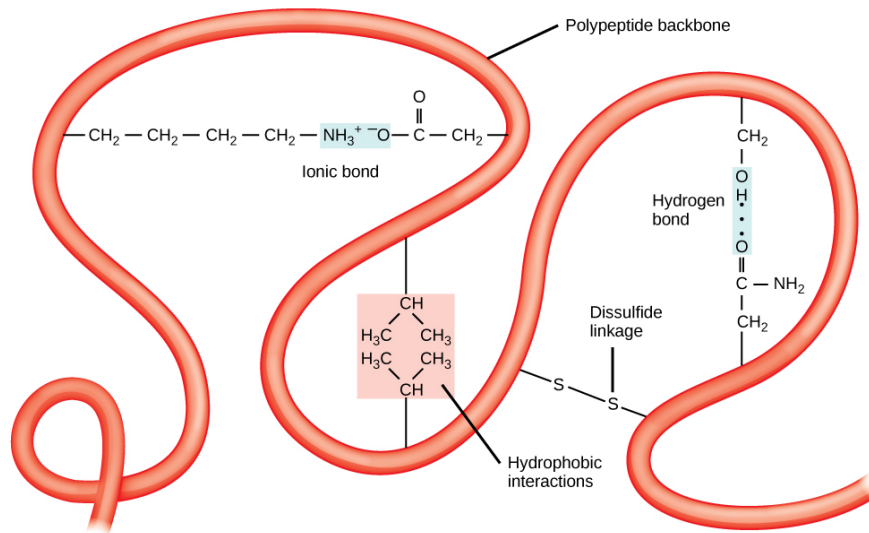


The α -helix and β -pleated sheet are secondary structures of proteins that form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an α -helix, while others have a propensity to form a β -pleated sheet.

Every helical turn in an alpha helix has 3.6 amino acid residues. The R groups (the variant groups) of the polypeptide protrude out from the α -helix chain. In the β -pleated sheet, the “pleats” are formed by hydrogen bonding between atoms on the backbone of the polypeptide chain. The R groups are attached to the carbons and extend above and below the folds of the pleat. The pleated segments align parallel or antiparallel to each other, and hydrogen bonds form between the partially positive nitrogen atom in the amino group and the partially negative oxygen atom in the carbonyl group of the peptide backbone. The α -helix and β -pleated sheet structures are found in most globular and fibrous proteins and they play an important structural role.

Tertiary Structure

The unique three-dimensional structure of a polypeptide is its **tertiary structure** ([\[link\]](#)). This structure is in part due to chemical interactions at work on the polypeptide chain. Primarily, the interactions among R groups creates the complex three-dimensional tertiary structure of a protein. The nature of the R groups found in the amino acids involved can counteract the formation of the hydrogen bonds described for standard secondary structures. For example, R groups with like charges are repelled by each other and those with unlike charges are attracted to each other (ionic bonds). When protein folding takes place, the hydrophobic R groups of nonpolar amino acids lay in the interior of the protein, whereas the hydrophilic R groups lay on the outside. The former types of interactions are also known as hydrophobic interactions. Interaction between cysteine side chains forms disulfide linkages in the presence of oxygen, the only covalent bond forming during protein folding.



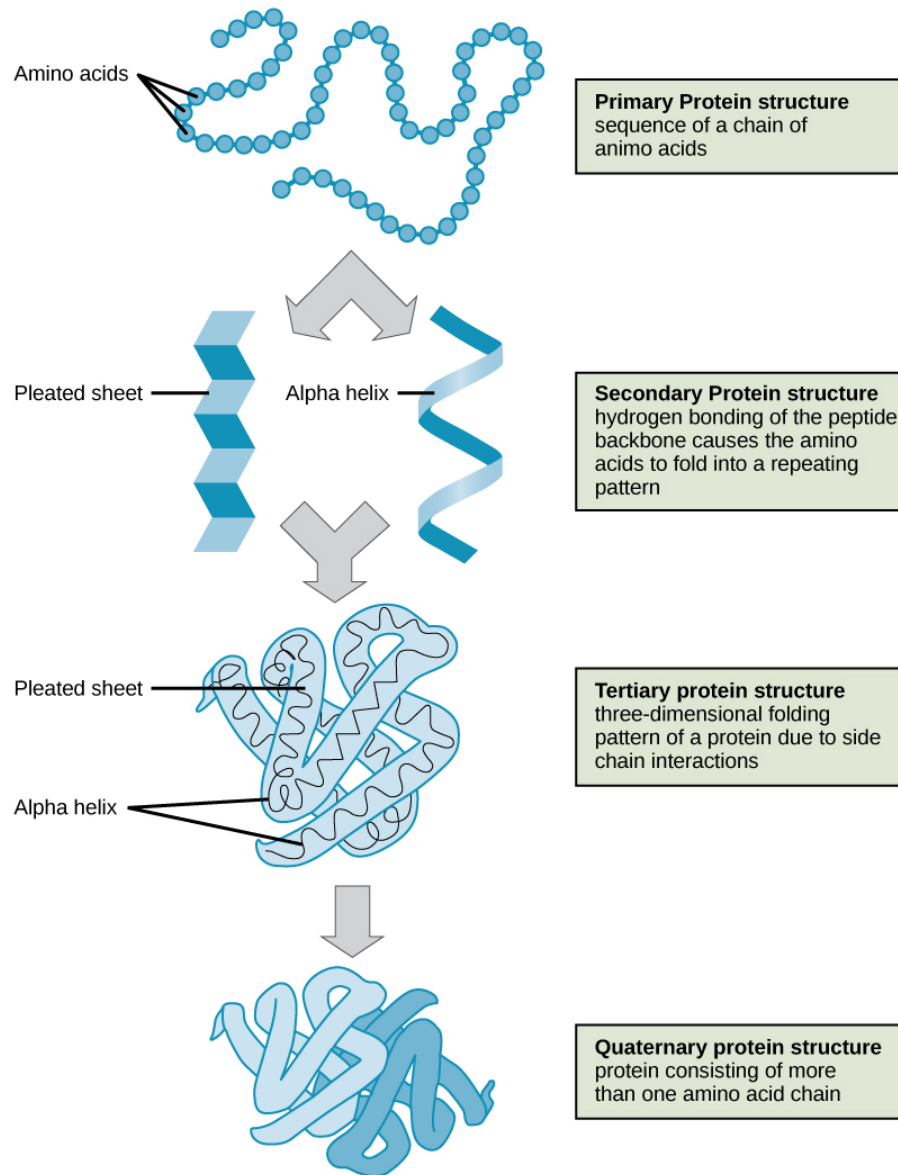
The tertiary structure of proteins is determined by a variety of chemical interactions. These include hydrophobic interactions, ionic bonding, hydrogen bonding and disulfide linkages.

All of these interactions, weak and strong, determine the final three-dimensional shape of the protein. When a protein loses its three-dimensional shape, it may no longer be functional.

Quaternary Structure

In nature, some proteins are formed from several polypeptides, also known as subunits, and the interaction of these subunits forms the **quaternary structure**. Weak interactions between the subunits help to stabilize the overall structure. For example, insulin (a globular protein) has a combination of hydrogen bonds and disulfide bonds that cause it to be mostly clumped into a ball shape. Insulin starts out as a single polypeptide and loses some internal sequences in the presence of post-translational modification after the formation of the disulfide linkages that hold the remaining chains together. Silk (a fibrous protein), however, has a β -pleated sheet structure that is the result of hydrogen bonding between different chains.

The four levels of protein structure (primary, secondary, tertiary, and quaternary) are illustrated in [\[link\]](#).



The four levels of protein structure can be observed in these illustrations. (credit: modification of work by National Human Genome Research Institute)

Denaturation and Protein Folding

Each protein has its own unique sequence and shape that are held together by chemical interactions. If the protein is subject to changes in temperature, pH, or exposure to chemicals, the protein structure may change, losing its shape without losing its primary sequence in what is known as denaturation. Denaturation is often reversible because the primary structure of the polypeptide is conserved in the process if the denaturing agent is removed, allowing the protein to resume its function. Sometimes denaturation is irreversible, leading to loss of function. One example of irreversible protein denaturation is when an egg is fried. The albumin protein in the liquid egg white is denatured when placed in a hot pan. Not all proteins are denatured at high temperatures; for instance, bacteria that survive in hot springs have proteins that function at temperatures close to boiling. The stomach is also very acidic, has a low pH, and denatures proteins as part of the digestion process; however, the digestive enzymes of the stomach retain their activity under these conditions.

Protein folding is critical to its function. It was originally thought that the proteins themselves were responsible for the folding process. Only recently was it found that often they receive assistance in the folding process from protein helpers known as **chaperones** (or chaperonins) that associate with the target protein during the folding process. They act by preventing aggregation of polypeptides that make up the complete protein structure, and they disassociate from the protein once the target protein is folded.

Note:

Link to Learning



For an additional perspective on proteins, view [this animation](#) called “Biomolecules: The Proteins.”

Section Summary

Proteins are a class of macromolecules that perform a diverse range of functions for the cell. They help in metabolism by providing structural support and by acting as enzymes, carriers, or hormones. The building blocks of proteins (monomers) are amino acids. Each amino acid has a central carbon that is linked to an amino group, a carboxyl group, a hydrogen atom, and an R group or side chain. There are 20 commonly occurring amino acids, each of which differs in the R group. Each amino acid is linked to its neighbors by a peptide bond. A long chain of amino acids is known as a polypeptide.

Proteins are organized at four levels: primary, secondary, tertiary, and (optional) quaternary. The primary structure is the unique sequence of amino acids. The local folding of the polypeptide to form structures such as the α helix and β -pleated sheet constitutes the secondary structure. The overall three-dimensional structure is the tertiary structure. When two or more polypeptides combine to form the complete protein structure, the configuration is known as the quaternary structure of a protein. Protein shape and function are intricately linked; any change in shape caused by changes in temperature or pH may lead to protein denaturation and a loss in function.

Art Connections

Exercise:

Problem:

[\[link\]](#) Which categories of amino acid would you expect to find on the surface of a soluble protein, and which would you expect to find in the interior? What distribution of amino acids would you expect to find in a protein embedded in a lipid bilayer?

Solution:

[\[link\]](#) Polar and charged amino acid residues (the remainder after peptide bond formation) are more likely to be found on the surface of

soluble proteins where they can interact with water, and nonpolar (e.g., amino acid side chains) are more likely to be found in the interior where they are sequestered from water. In membrane proteins, nonpolar and hydrophobic amino acid side chains associate with the hydrophobic tails of phospholipids, while polar and charged amino acid side chains interact with the polar head groups or with the aqueous solution. However, there are exceptions. Sometimes, positively and negatively charged amino acid side chains interact with one another in the interior of a protein, and polar or charged amino acid side chains that interact with a ligand can be found in the ligand binding pocket.

Review Questions

Exercise:

Problem: The monomers that make up proteins are called _____.

- a. nucleotides
- b. disaccharides
- c. amino acids
- d. chaperones

Solution:

C

Exercise:

Problem:

The α helix and the β -pleated sheet are part of which protein structure?

- a. primary
- b. secondary
- c. tertiary
- d. quaternary

Solution:

B

Free Response**Exercise:****Problem:**

Explain what happens if even one amino acid is substituted for another in a polypeptide chain. Provide a specific example.

Solution:

A change in gene sequence can lead to a different amino acid being added to a polypeptide chain instead of the normal one. This causes a change in protein structure and function. For example, in sickle cell anemia, the hemoglobin β chain has a single amino acid substitution—the amino acid glutamic acid in position six is substituted by valine. Because of this change, hemoglobin molecules form aggregates, and the disc-shaped red blood cells assume a crescent shape, which results in serious health problems.

Exercise:

Problem: Describe the differences in the four protein structures.

Solution:

The sequence and number of amino acids in a polypeptide chain is its primary structure. The local folding of the polypeptide in some regions is the secondary structure of the protein. The three-dimensional structure of a polypeptide is known as its tertiary structure, created in part by chemical interactions such as hydrogen bonds between polar side chains, van der Waals interactions, disulfide linkages, and

hydrophobic interactions. Some proteins are formed from multiple polypeptides, also known as subunits, and the interaction of these subunits forms the quaternary structure.

Glossary

alpha-helix structure (α -helix)

type of secondary structure of proteins formed by folding of the polypeptide into a helix shape with hydrogen bonds stabilizing the structure

amino acid

monomer of a protein; has a central carbon or alpha carbon to which an amino group, a carboxyl group, a hydrogen, and an R group or side chain is attached; the R group is different for all 20 amino acids

beta-pleated sheet (β -pleated)

secondary structure found in proteins in which “pleats” are formed by hydrogen bonding between atoms on the backbone of the polypeptide chain

chaperone

(also, chaperonin) protein that helps nascent protein in the folding process

denaturation

loss of shape in a protein as a result of changes in temperature, pH, or exposure to chemicals

enzyme

catalyst in a biochemical reaction that is usually a complex or conjugated protein

hormone

chemical signaling molecule, usually protein or steroid, secreted by endocrine cells that act to control or regulate specific physiological processes

peptide bond

bond formed between two amino acids by a dehydration reaction

polypeptide

long chain of amino acids linked by peptide bonds

primary structure

linear sequence of amino acids in a protein

protein

biological macromolecule composed of one or more chains of amino acids

quaternary structure

association of discrete polypeptide subunits in a protein

secondary structure

regular structure formed by proteins by intramolecular hydrogen bonding between the oxygen atom of one amino acid residue and the hydrogen attached to the nitrogen atom of another amino acid residue

tertiary structure

three-dimensional conformation of a protein, including interactions between secondary structural elements; formed from interactions between amino acid side chains

Nucleic Acids

By the end of this section, you will be able to:

- Describe the structure of nucleic acids and define the two types of nucleic acids
- Explain the structure and role of DNA
- Explain the structure and roles of RNA

Nucleic acids are the most important macromolecules for the continuity of life. They carry the genetic blueprint of a cell and carry instructions for the functioning of the cell.

DNA and RNA

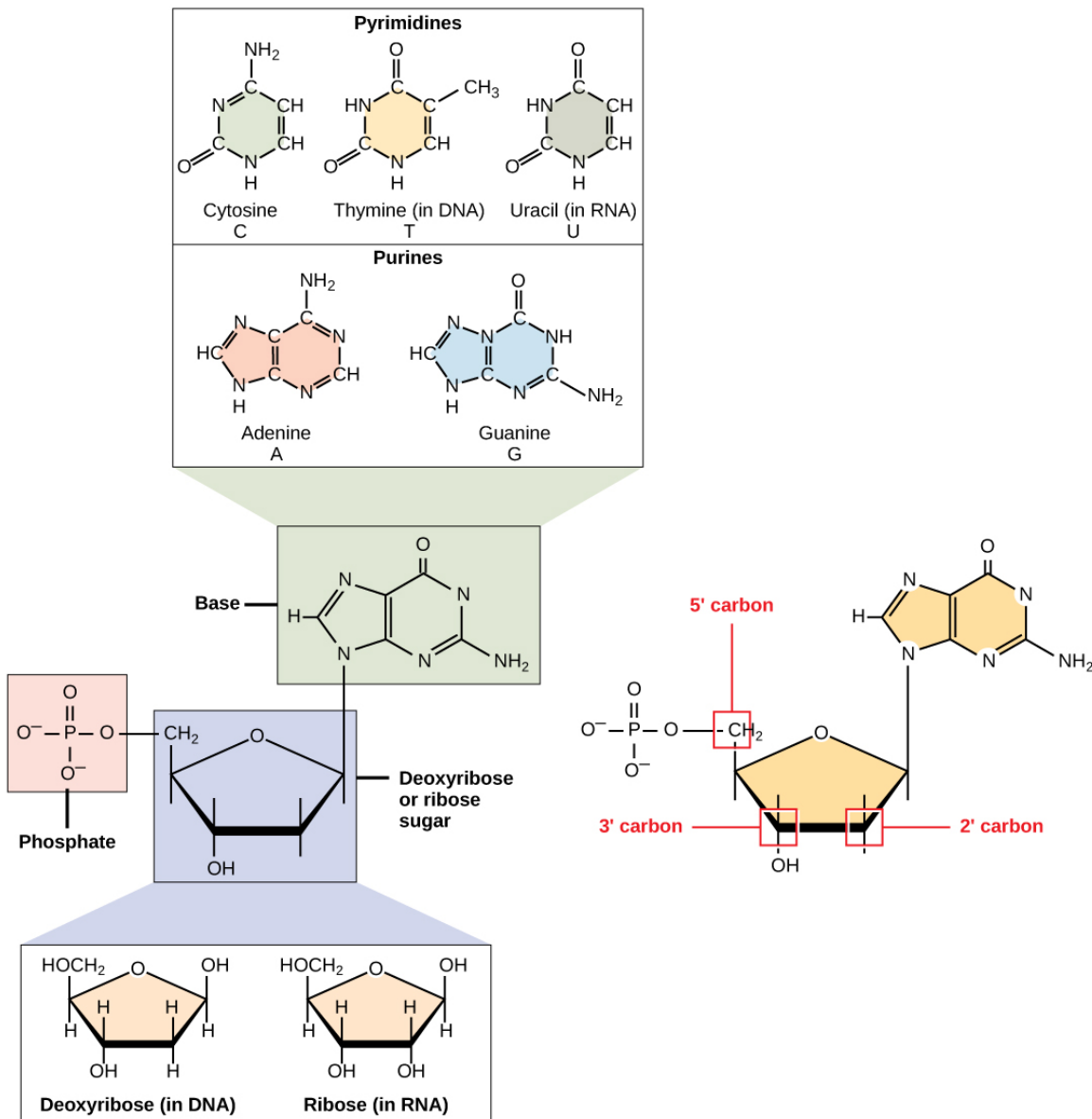
The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms, ranging from single-celled bacteria to multicellular mammals. It is found in the nucleus of eukaryotes and in the organelles, chloroplasts, and mitochondria. In prokaryotes, the DNA is not enclosed in a membranous envelope.

The entire genetic content of a cell is known as its genome, and the study of genomes is genomics. In eukaryotic cells but not in prokaryotes, DNA forms a complex with histone proteins to form chromatin, the substance of eukaryotic chromosomes. A chromosome may contain tens of thousands of genes. Many genes contain the information to make protein products; other genes code for RNA products. DNA controls all of the cellular activities by turning the genes “on” or “off.”

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus but instead use an intermediary to communicate with the rest of the cell. This intermediary is the **messenger RNA (mRNA)**. Other types of RNA—like rRNA, tRNA, and microRNA—are involved in protein synthesis and its regulation.

DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a **polynucleotide**, DNA or

RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group ([\[link\]](#)). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to one or more phosphate groups.



A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbon residues in the pentose are numbered 1' through 5' (the prime distinguishes these residues from those in the base, which are numbered without using a

prime notation). The base is attached to the 1' position of the ribose, and the phosphate is attached to the 5' position. When a polynucleotide is formed, the 5' phosphate of the incoming nucleotide attaches to the 3' hydroxyl group at the end of the growing chain. Two types of pentose are found in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an H instead of an OH at the 2' position. Bases can be divided into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring.

The nitrogenous bases, important components of nucleotides, are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus, decreases the hydrogen ion concentration in its environment, making it more basic. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T).

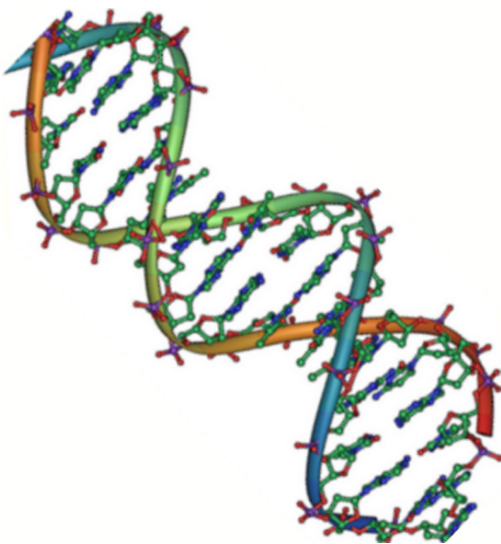
Adenine and guanine are classified as **purines**. The primary structure of a purine is two carbon-nitrogen rings. Cytosine, thymine, and uracil are classified as **pyrimidines** which have a single carbon-nitrogen ring as their primary structure ([\[link\]](#)). Each of these basic carbon-nitrogen rings has different functional groups attached to it. In molecular biology shorthand, the nitrogenous bases are simply known by their symbols A, T, G, C, and U. DNA contains A, T, G, and C whereas RNA contains A, U, G, and C.

The pentose sugar in DNA is deoxyribose, and in RNA, the sugar is ribose ([\[link\]](#)). The difference between the sugars is the presence of the hydroxyl group on the second carbon of the ribose and hydrogen on the second carbon of the deoxyribose. The carbon atoms of the sugar molecule are numbered as 1', 2', 3', 4', and 5' (1' is read as "one prime"). The phosphate residue is attached to the hydroxyl group of the 5' carbon of one sugar and the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, which forms a 5'–3' **phosphodiester** linkage. The phosphodiester linkage is not formed by simple dehydration reaction like the other linkages

connecting monomers in macromolecules: its formation involves the removal of two phosphate groups. A polynucleotide may have thousands of such phosphodiester linkages.

DNA Double-Helix Structure

DNA has a double-helix structure ([\[link\]](#)). The sugar and phosphate lie on the outside of the helix, forming the backbone of the DNA. The nitrogenous bases are stacked in the interior, like the steps of a staircase, in pairs; the pairs are bound to each other by hydrogen bonds. Every base pair in the double helix is separated from the next base pair by 0.34 nm. The two strands of the helix run in opposite directions, meaning that the 5' carbon end of one strand will face the 3' carbon end of its matching strand. (This is referred to as antiparallel orientation and is important to DNA replication and in many nucleic acid interactions.)



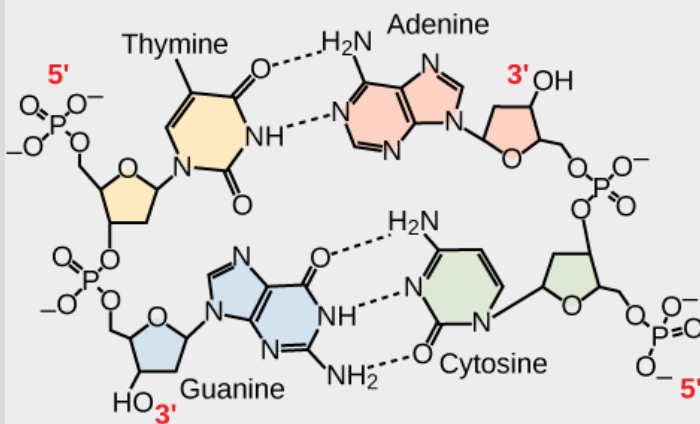
Native DNA is an antiparallel double helix. The phosphate backbone (indicated by the curvy lines) is on the outside, and the bases are on the

inside. Each base from one strand interacts via hydrogen bonding with a base from the opposing strand. (credit: Jerome Walker/Dennis Myts)

Only certain types of base pairing are allowed. For example, a certain purine can only pair with a certain pyrimidine. This means A can pair with T, and G can pair with C, as shown in [\[link\]](#). This is known as the base complementary rule. In other words, the DNA strands are complementary to each other. If the sequence of one strand is AATTGGCC, the complementary strand would have the sequence TTAACCGG. During DNA replication, each strand is copied, resulting in a daughter DNA double helix containing one parental DNA strand and a newly synthesized strand.

Note:

Art Connection



In a double stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5' to 3' and the other 3' to 5'. The phosphate backbone is located on the outside,

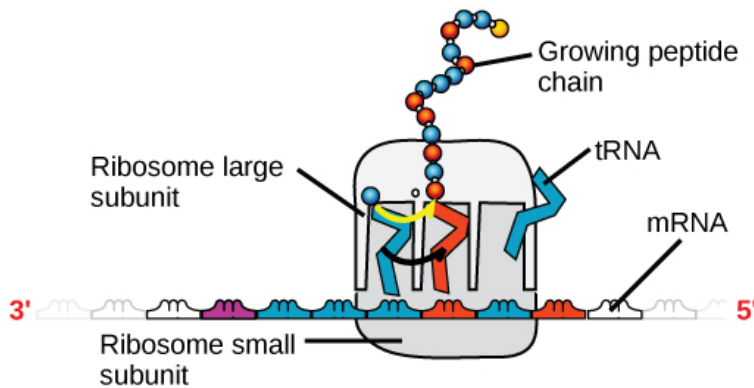
and the bases are in the middle.
Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine.

A mutation occurs, and cytosine is replaced with adenine. What impact do you think this will have on the DNA structure?

RNA

Ribonucleic acid, or RNA, is mainly involved in the process of protein synthesis under the direction of DNA. RNA is usually single-stranded and is made of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and the phosphate group.

There are four major types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and microRNA (miRNA). The first, mRNA, carries the message from DNA, which controls all of the cellular activities in a cell. If a cell requires a certain protein to be synthesized, the gene for this product is turned “on” and the messenger RNA is synthesized in the nucleus. The RNA base sequence is complementary to the coding sequence of the DNA from which it has been copied. However, in RNA, the base T is absent and U is present instead. If the DNA strand has a sequence AATTGCGC, the sequence of the complementary RNA is UUAACGCG. In the cytoplasm, the mRNA interacts with ribosomes and other cellular machinery ([link](#)).



A ribosome has two parts: a large subunit and a small subunit. The mRNA sits in between the two subunits. A tRNA molecule recognizes a codon on the mRNA, binds to it by complementary base pairing, and adds the correct amino acid to the growing peptide chain.

The mRNA is read in sets of three bases known as codons. Each codon codes for a single amino acid. In this way, the mRNA is read and the protein product is made. **Ribosomal RNA (rRNA)** is a major constituent of ribosomes on which the mRNA binds. The rRNA ensures the proper alignment of the mRNA and the ribosomes; the rRNA of the ribosome also has an enzymatic activity (peptidyl transferase) and catalyzes the formation of the peptide bonds between two aligned amino acids. **Transfer RNA (tRNA)** is one of the smallest of the four types of RNA, usually 70–90 nucleotides long. It carries the correct amino acid to the site of protein synthesis. It is the base pairing between the tRNA and mRNA that allows for the correct amino acid to be inserted in the polypeptide chain. microRNAs are the smallest RNA molecules and their role involves the regulation of gene expression by interfering with the expression of certain mRNA messages. [\[link\]](#) summarizes features of DNA and RNA.

Features of DNA and RNA		
	DNA	RNA
Function	Carries genetic information	Involved in protein synthesis
Location	Remains in the nucleus	Leaves the nucleus
Structure	Double helix	Usually single-stranded
Sugar	Deoxyribose	Ribose
Pyrimidines	Cytosine, thymine	Cytosine, uracil
Purines	Adenine, guanine	Adenine, guanine

Even though the RNA is single stranded, most RNA types show extensive intramolecular base pairing between complementary sequences, creating a predictable three-dimensional structure essential for their function.

As you have learned, information flow in an organism takes place from DNA to RNA to protein. DNA dictates the structure of mRNA in a process known as **transcription**, and RNA dictates the structure of protein in a process known as **translation**. This is known as the Central Dogma of Life, which holds true for all organisms; however, exceptions to the rule occur in connection with viral infections.

Note:

Link to Learning



To learn more about DNA, explore the [Howard Hughes Medical Institute BioInteractive animations](#) on the topic of DNA.

Section Summary

Nucleic acids are molecules made up of nucleotides that direct cellular activities such as cell division and protein synthesis. Each nucleotide is made up of a pentose sugar, a nitrogenous base, and a phosphate group. There are two types of nucleic acids: DNA and RNA. DNA carries the genetic blueprint of the cell and is passed on from parents to offspring (in the form of chromosomes). It has a double-helical structure with the two strands running in opposite directions, connected by hydrogen bonds, and complementary to each other. RNA is single-stranded and is made of a pentose sugar (ribose), a nitrogenous base, and a phosphate group. RNA is involved in protein synthesis and its regulation. Messenger RNA (mRNA) is copied from the DNA, is exported from the nucleus to the cytoplasm, and contains information for the construction of proteins. Ribosomal RNA (rRNA) is a part of the ribosomes at the site of protein synthesis, whereas transfer RNA (tRNA) carries the amino acid to the site of protein synthesis. microRNA regulates the use of mRNA for protein synthesis.

Art Connections

Exercise:

Problem:

[\[link\]](#) A mutation occurs, and cytosine is replaced with adenine. What impact do you think this will have on the DNA structure?

Solution:

[\[link\]](#) Adenine is larger than cytosine and will not be able to base pair properly with the guanine on the opposing strand. This will cause the DNA to bulge. DNA repair enzymes may recognize the bulge and replace the incorrect nucleotide.

Review Questions**Exercise:**

Problem: A nucleotide of DNA may contain _____.

- a. ribose, uracil, and a phosphate group
- b. deoxyribose, uracil, and a phosphate group
- c. deoxyribose, thymine, and a phosphate group
- d. ribose, thymine, and a phosphate group

Solution:

C

Exercise:

Problem: The building blocks of nucleic acids are _____.

- a. sugars
- b. nitrogenous bases
- c. peptides
- d. nucleotides

Solution:

D

Free Response

Exercise:

Problem:What are the structural differences between RNA and DNA?

Solution:

DNA has a double-helix structure. The sugar and the phosphate are on the outside of the helix and the nitrogenous bases are in the interior. The monomers of DNA are nucleotides containing deoxyribose, one of the four nitrogenous bases (A, T, G and C), and a phosphate group. RNA is usually single-stranded and is made of ribonucleotides that are linked by phosphodiester linkages. A ribonucleotide contains ribose (the pentose sugar), one of the four nitrogenous bases (A,U, G, and C), and the phosphate group.

Exercise:

Problem:What are the four types of RNA and how do they function?

Solution:

The four types of RNA are messenger RNA, ribosomal RNA, transfer RNA, and microRNA. Messenger RNA carries the information from the DNA that controls all cellular activities. The mRNA binds to the ribosomes that are constructed of proteins and rRNA, and tRNA transfers the correct amino acid to the site of protein synthesis. microRNA regulates the availability of mRNA for translation.

Glossary

deoxyribonucleic acid (DNA)

double-helical molecule that carries the hereditary information of the cell

messenger RNA (mRNA)

RNA that carries information from DNA to ribosomes during protein synthesis

nucleic acid

biological macromolecule that carries the genetic blueprint of a cell and carries instructions for the functioning of the cell

nucleotide

monomer of nucleic acids; contains a pentose sugar, one or more phosphate groups, and a nitrogenous base

phosphodiester

linkage covalent chemical bond that holds together the polynucleotide chains with a phosphate group linking two pentose sugars of neighboring nucleotides

polynucleotide

long chain of nucleotides

purine

type of nitrogenous base in DNA and RNA; adenine and guanine are purines

pyrimidine

type of nitrogenous base in DNA and RNA; cytosine, thymine, and uracil are pyrimidines

ribonucleic acid (RNA)

single-stranded, often internally base paired, molecule that is involved in protein synthesis

ribosomal RNA (rRNA)

RNA that ensures the proper alignment of the mRNA and the ribosomes during protein synthesis and catalyzes the formation of the peptide linkage

transcription

process through which messenger RNA forms on a template of DNA

transfer RNA (tRNA)

RNA that carries activated amino acids to the site of protein synthesis on the ribosome

translation

process through which RNA directs the formation of protein

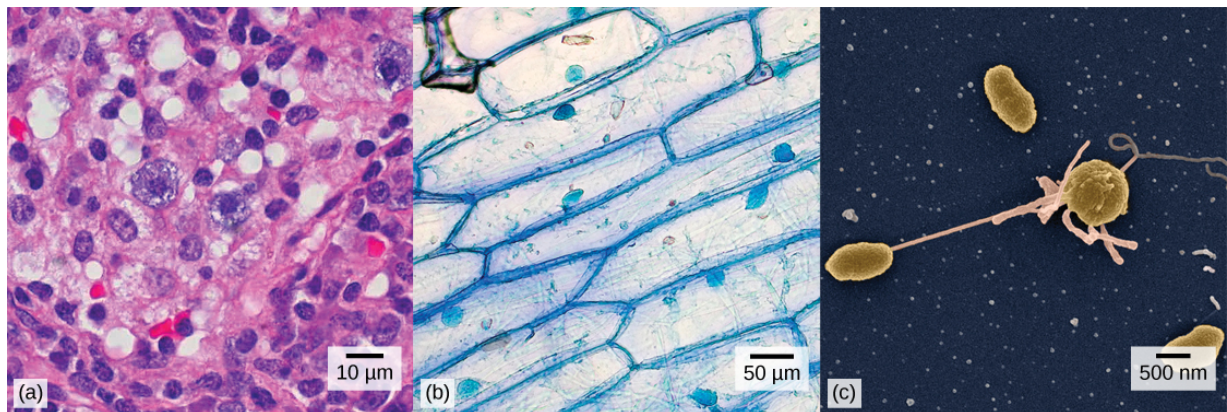
Introduction

class="introduction"

(a) Nasal
sinus cells
(viewed with
a light
microscope),
(b) onion
cells (viewed
with a light
microscope),
and (c) *Vibrio
tasmaniensis*
bacterial cells
(seen through
a scanning
electron
microscope)
are from very
different
organisms,
yet all share
certain
characteristic
s of basic cell
structure.

(credit a:
modification
of work by
Ed Uthman,
MD; credit b:
modification
of work by
Umberto
Salvagnin;
credit c:

modification
of work by
Anthony
D'Onofrio,
William H.
Fowle, Eric J.
Stewart, and
Kim Lewis of
the Lewis
Lab at
Northeastern
University;
scale-bar data
from Matt
Russell)



Close your eyes and picture a brick wall. What is the basic building block of that wall? A single brick, of course. Like a brick wall, your body is composed of basic building blocks, and the building blocks of your body are cells.

Your body has many kinds of cells, each specialized for a specific purpose. Just as a home is made from a variety of building materials, the human body is constructed from many cell types. For example, epithelial cells protect the surface of the body and cover the organs and body cavities within. Bone cells help to support and protect the body. Cells of the immune

system fight invading bacteria. Additionally, blood and blood cells carry nutrients and oxygen throughout the body while removing carbon dioxide. Each of these cell types plays a vital role during the growth, development, and day-to-day maintenance of the body. In spite of their enormous variety, however, cells from all organisms—even ones as diverse as bacteria, onion, and human—share certain fundamental characteristics.

Studying Cells

By the end of this section, you will be able to:

- Describe the role of cells in organisms
- Compare and contrast light microscopy and electron microscopy
- Summarize cell theory

A cell is the smallest unit of a living thing. A living thing, whether made of one cell (like bacteria) or many cells (like a human), is called an organism. Thus, cells are the basic building blocks of all organisms.

Several cells of one kind that interconnect with each other and perform a shared function form tissues, several tissues combine to form an organ (your stomach, heart, or brain), and several organs make up an organ system (such as the digestive system, circulatory system, or nervous system). Several systems that function together form an organism (like a human being). Here, we will examine the structure and function of cells.

There are many types of cells, all grouped into one of two broad categories: prokaryotic and eukaryotic. For example, both animal and plant cells are classified as eukaryotic cells, whereas bacterial cells are classified as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, let's first examine how biologists study cells.

Microscopy

Cells vary in size. With few exceptions, individual cells cannot be seen with the naked eye, so scientists use microscopes (micro- = “small”; -scope = “to look at”) to study them. A **microscope** is an instrument that magnifies an object. Most photographs of cells are taken with a microscope, and these images can also be called micrographs.

The optics of a microscope's lenses change the orientation of the image that the user sees. A specimen that is right-side up and facing right on the microscope slide will appear upside-down and facing left when viewed through a microscope, and vice versa. Similarly, if the slide is moved left while looking through the microscope, it will appear to move right, and if

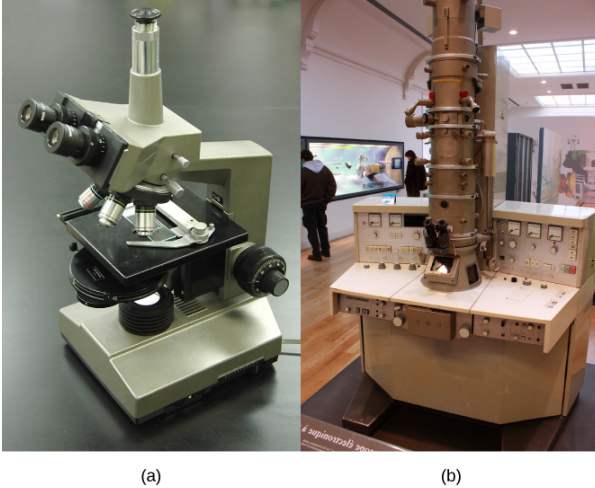
moved down, it will seem to move up. This occurs because microscopes use two sets of lenses to magnify the image. Because of the manner by which light travels through the lenses, this system of two lenses produces an inverted image (binocular, or dissecting microscopes, work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

Light Microscopes

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter or eight micrometers (abbreviated as eight μm) in diameter; the head of a pin is about two thousandths of a meter (two mm) in diameter. That means about 250 red blood cells could fit on the head of a pin.

Most student microscopes are classified as **light microscopes** ([link](#)a). Visible light passes and is bent through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains. Staining, however, usually kills the cells.

Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. When oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times. In order to gain a better understanding of cellular structure and function, scientists typically use electron microscopes.



(a) Most light microscopes used in a college biology lab can magnify cells up to approximately 400 times and have a resolution of about 200 nanometers. (b) Electron microscopes provide a much higher magnification, 100,000x, and have a resolution of 50 picometers. (credit a: modification of work by "GcG"/Wikimedia Commons; credit b: modification of work by Evan Bench)

Electron Microscopes

In contrast to light microscopes, **electron microscopes** ([link](#)) use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail ([link](#)), it also provides higher resolving power. The method used to prepare the specimen for viewing with an electron microscope kills the specimen. Electrons have short

wavelengths (shorter than photons) that move best in a vacuum, so living cells cannot be viewed with an electron microscope.

In a scanning electron microscope, a beam of electrons moves back and forth across a cell's surface, creating details of cell surface characteristics. In a transmission electron microscope, the electron beam penetrates the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than light microscopes.



(a) These *Salmonella* bacteria appear as tiny purple dots when viewed with a light microscope. (b) This scanning electron microscope micrograph shows *Salmonella* bacteria (in red) invading human cells (yellow). Even though subfigure (b) shows a different *Salmonella* specimen than subfigure (a), you can still observe the comparative increase in magnification and detail. (credit a: modification of work by CDC/Armed Forces Institute of Pathology, Charles N. Farmer, Rocky Mountain Laboratories; credit b: modification of work by NIAID, NIH; scale-bar data from Matt Russell)

Note:

Link to Learning



For another perspective on cell size, try the HowBig interactive at [this site](#).

Cell Theory

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of protista (a type of single-celled organism) and sperm, which he collectively termed “animalcules.”

In a 1665 publication called *Micrographia*, experimental scientist Robert Hooke coined the term “cell” for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses, microscope construction, and staining techniques enabled other scientists to see some components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the **unified cell theory**, which states that all living things are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells. Rudolf Virchow later made important contributions to this theory.

Note:

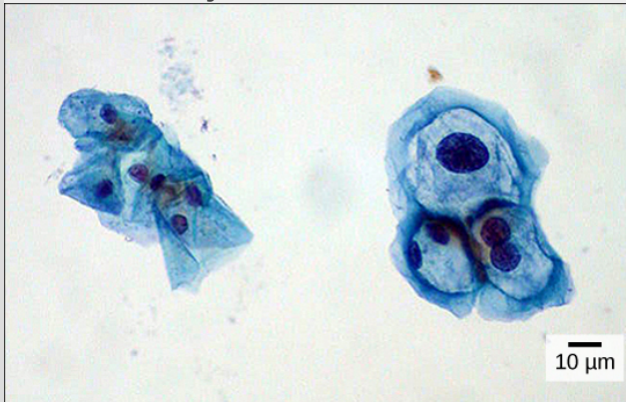
Career Connection

Cytotechnologist

Have you ever heard of a medical test called a Pap smear ([link](#))? In this test, a doctor takes a small sample of cells from the uterine cervix of a patient and sends it to a medical lab where a cytotechnologist stains the cells and examines them for any changes that could indicate cervical cancer or a microbial infection.

Cytotechnologists (cyto- = “cell”) are professionals who study cells via microscopic examinations and other laboratory tests. They are trained to determine which cellular changes are within normal limits and which are abnormal. Their focus is not limited to cervical cells; they study cellular specimens that come from all organs. When they notice abnormalities, they consult a pathologist, who is a medical doctor who can make a clinical diagnosis.

Cytotechnologists play a vital role in saving people’s lives. When abnormalities are discovered early, a patient’s treatment can begin sooner, which usually increases the chances of a successful outcome.



These uterine cervix cells, viewed through a light microscope, were obtained from a Pap smear.

Normal cells are on the left. The cells on the right are infected with human papillomavirus (HPV).

Notice that the infected cells are larger; also, two of these cells each have two nuclei instead of one, the normal number. (credit:

modification of work by Ed
Uthman, MD; scale-bar data from
Matt Russell)

Section Summary

A cell is the smallest unit of life. Most cells are so tiny that they cannot be seen with the naked eye. Therefore, scientists use microscopes to study cells. Electron microscopes provide higher magnification, higher resolution, and more detail than light microscopes. The unified cell theory states that all organisms are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells.

Review Questions

Exercise:

Problem:

When viewing a specimen through a light microscope, scientists use _____ to distinguish the individual components of cells.

- a. a beam of electrons
- b. radioactive isotopes
- c. special stains
- d. high temperatures

Solution:

C

Exercise:

Problem: The _____ is the basic unit of life.

- a. organism
- b. cell
- c. tissue
- d. organ

Solution:

B

Free Response

Exercise:

Problem:

In your everyday life, you have probably noticed that certain instruments are ideal for certain situations. For example, you would use a spoon rather than a fork to eat soup because a spoon is shaped for scooping, while soup would slip between the tines of a fork. The use of ideal instruments also applies in science. In what situation(s) would the use of a light microscope be ideal, and why?

Solution:

A light microscope would be ideal when viewing a small living organism, especially when the cell has been stained to reveal details.

Exercise:

Problem:

In what situation(s) would the use of a scanning electron microscope be ideal, and why?

Solution:

A scanning electron microscope would be ideal when you want to view the minute details of a cell's surface, because its beam of electrons moves back and forth over the surface to convey the image.

Exercise:**Problem:**

In what situation(s) would a transmission electron microscope be ideal, and why?

Solution:

A transmission electron microscope would be ideal for viewing the cell's internal structures, because many of the internal structures have membranes that are not visible by the light microscope.

Exercise:**Problem:**

What are the advantages and disadvantages of each of these types of microscopes?

Solution:

The advantages of light microscopes are that they are easily obtained, and the light beam does not kill the cells. However, typical light microscopes are somewhat limited in the amount of detail they can reveal. Electron microscopes are ideal because you can view intricate details, but they are bulky and costly, and preparation for the microscopic examination kills the specimen.

Glossary

cell theory
 see unified cell theory

electron microscope

an instrument that magnifies an object using a beam of electrons passed and bent through a lens system to visualize a specimen

light microscope

an instrument that magnifies an object using a beam visible light passed and bent through a lens system to visualize a specimen

microscope

an instrument that magnifies an object

unified cell theory

a biological concept that states that all organisms are composed of one or more cells; the cell is the basic unit of life; and new cells arise from existing cells

Prokaryotic Cells

By the end of this section, you will be able to:

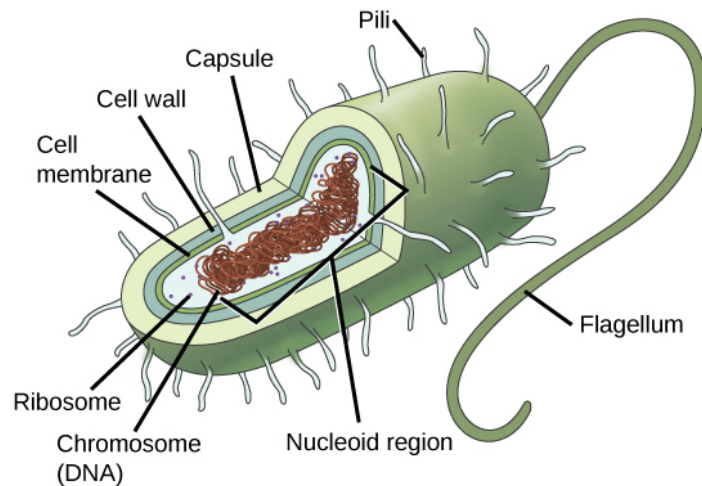
- Name examples of prokaryotic and eukaryotic organisms
- Compare and contrast prokaryotic cells and eukaryotic cells
- Describe the relative sizes of different kinds of cells
- Explain why cells must be small

Cells fall into one of two broad categories: prokaryotic and eukaryotic. Only the predominantly single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes (pro- = “before”; -kary- = “nucleus”). Animals, plants, fungi, and protists are all eukaryotes (eu- = “true”) and are made up of eukaryotic cells.

Components of Prokaryotic Cells

All cells share four common components: 1) a plasma membrane, an outer covering that separates the cell’s interior from its surrounding environment; 2) cytoplasm, consisting of a jelly-like cytosol within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, which synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryote** is a simple, mostly single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in a central part of the cell: the **nucleoid** ([\[link\]](#)).



This figure shows the generalized structure of a prokaryotic cell. All prokaryotes have chromosomal DNA localized in a nucleoid, ribosomes, a cell membrane, and a cell wall. The other structures shown are present in some, but not all, bacteria.

Most prokaryotes have a peptidoglycan cell wall and many have a polysaccharide capsule ([link](#)). The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion. Pili are used to exchange genetic material during a type of reproduction called conjugation. Fimbriae are used by bacteria to attach to a host cell.

Note:**Career Connection****Microbiologist**

The most effective action anyone can take to prevent the spread of contagious illnesses is to wash his or her hands. Why? Because microbes

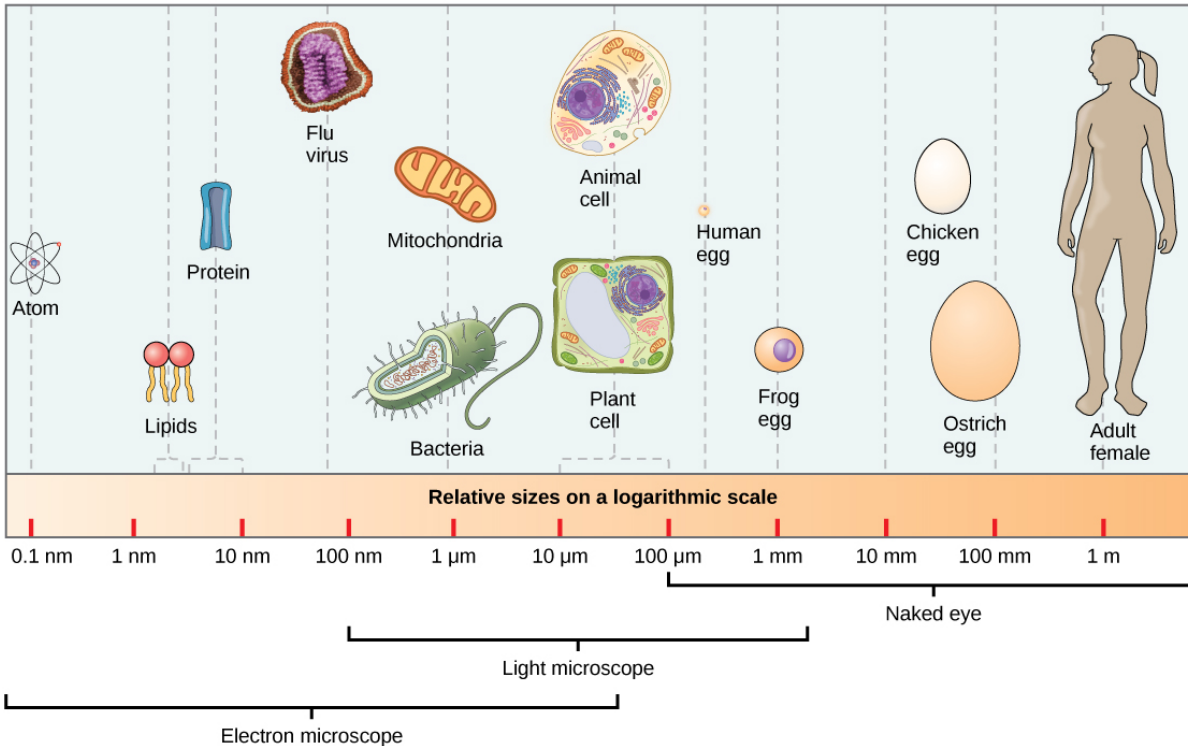
(organisms so tiny that they can only be seen with microscopes) are ubiquitous. They live on doorknobs, money, your hands, and many other surfaces. If someone sneezes into his hand and touches a doorknob, and afterwards you touch that same doorknob, the microbes from the sneezer's mucus are now on your hands. If you touch your hands to your mouth, nose, or eyes, those microbes can enter your body and could make you sick.

However, not all microbes (also called microorganisms) cause disease; most are actually beneficial. You have microbes in your gut that make vitamin K. Other microorganisms are used to ferment beer and wine. Microbiologists are scientists who study microbes. Microbiologists can pursue a number of careers. Not only do they work in the food industry, they are also employed in the veterinary and medical fields. They can work in the pharmaceutical sector, serving key roles in research and development by identifying new sources of antibiotics that could be used to treat bacterial infections.

Environmental microbiologists may look for new ways to use specially selected or genetically engineered microbes for the removal of pollutants from soil or groundwater, as well as hazardous elements from contaminated sites. These uses of microbes are called bioremediation technologies. Microbiologists can also work in the field of bioinformatics, providing specialized knowledge and insight for the design, development, and specificity of computer models of, for example, bacterial epidemics.

Cell Size

At 0.1 to 5.0 μm in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10 to 100 μm ([link](#)). The small size of prokaryotes allows ions and organic molecules that enter them to quickly diffuse to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly diffuse out. This is not the case in eukaryotic cells, which have developed different structural adaptations to enhance intracellular transport.



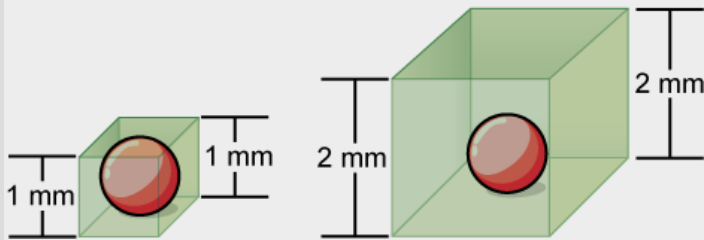
This figure shows relative sizes of microbes on a logarithmic scale (recall that each unit of increase in a logarithmic scale represents a 10-fold increase in the quantity being measured).

Small size, in general, is necessary for all cells, whether prokaryotic or eukaryotic. Let's examine why that is so. First, we'll consider the area and volume of a typical cell. Not all cells are spherical in shape, but most tend to approximate a sphere. You may remember from your high school geometry course that the formula for the surface area of a sphere is $4\pi r^2$, while the formula for its volume is $\frac{4\pi r^3}{3}$. Thus, as the radius of a cell increases, its surface area increases as the square of its radius, but its volume increases as the cube of its radius (much more rapidly). Therefore, as a cell increases in size, its surface area-to-volume ratio decreases. This same principle would apply if the cell had the shape of a cube ([\[link\]](#)). If the cell grows too large, the plasma membrane will not have sufficient surface area to support the rate of diffusion required for the increased volume. In other words, as a cell grows, it becomes less efficient. One way to become more efficient is to divide; another way is to develop organelles that

perform specific tasks. These adaptations lead to the development of more sophisticated cells called eukaryotic cells.

Note:

Art Connection



Notice that as a cell increases in size, its surface area-to-volume ratio decreases. When there is insufficient surface area to support a cell's increasing volume, a cell will either divide or die. The cell on the left has a volume of 1 mm^3 and a surface area of 6 mm^2 , with a surface area-to-volume ratio of 6 to 1, whereas the cell on the right has a volume of 8 mm^3 and a surface area of 24 mm^2 , with a surface area-to-volume ratio of 3 to 1.

Prokaryotic cells are much smaller than eukaryotic cells. What advantages might small cell size confer on a cell? What advantages might large cell size have?

Section Summary

Prokaryotes are predominantly single-celled organisms of the domains Bacteria and Archaea. All prokaryotes have plasma membranes, cytoplasm, ribosomes, and DNA that is not membrane-bound. Most have peptidoglycan cell walls and many have polysaccharide capsules. Prokaryotic cells range in diameter from 0.1 to 5.0 μm .

As a cell increases in size, its surface area-to-volume ratio decreases. If the cell grows too large, the plasma membrane will not have sufficient surface area to support the rate of diffusion required for the increased volume.

Art Connections

Exercise:

Problem:

[\[link\]](#) Prokaryotic cells are much smaller than eukaryotic cells. What advantages might small cell size confer on a cell? What advantages might large cell size have?

Solution:

[\[link\]](#) Substances can diffuse more quickly through small cells. Small cells have no need for organelles and therefore do not need to expend energy getting substances across organelle membranes. Large cells have organelles that can separate cellular processes, enabling them to build molecules that are more complex.

Review Questions

Exercise:

Problem:

Prokaryotes depend on _____ to obtain some materials and to get rid of wastes.

- a. ribosomes

- b. flagella
- c. cell division
- d. diffusion

Solution:

D

Exercise:

Problem: Bacteria that lack fimbriae are less likely to _____.

- a. adhere to cell surfaces
- b. swim through bodily fluids
- c. synthesize proteins
- d. retain the ability to divide

Solution:

A

Free Response

Exercise:

Problem:

Antibiotics are medicines that are used to fight bacterial infections. These medicines kill prokaryotic cells without harming human cells. What part or parts of the bacterial cell do you think antibiotics target? Why?

Solution:

The cell wall would be targeted by antibiotics as well as the bacteria's ability to replicate. This would inhibit the bacteria's ability to

reproduce, and it would compromise its defense mechanisms.

Exercise:

Problem: Explain why not all microbes are harmful.

Solution:

Some microbes are beneficial. For instance, *E. coli* bacteria populate the human gut and help break down fiber in the diet. Some foods such as yogurt are formed by bacteria.

Glossary

nucleoid

central part of a prokaryotic cell in which the chromosome is found

prokaryote

unicellular organism that lacks a nucleus or any other membrane-bound organelle

Eukaryotic Cells

By the end of this section, you will be able to:

- Describe the structure of eukaryotic cells
- Compare animal cells with plant cells
- State the role of the plasma membrane
- Summarize the functions of the major cell organelles

Have you ever heard the phrase “form follows function?” It’s a philosophy practiced in many industries. In architecture, this means that buildings should be constructed to support the activities that will be carried out inside them. For example, a skyscraper should be built with several elevator banks; a hospital should be built so that its emergency room is easily accessible.

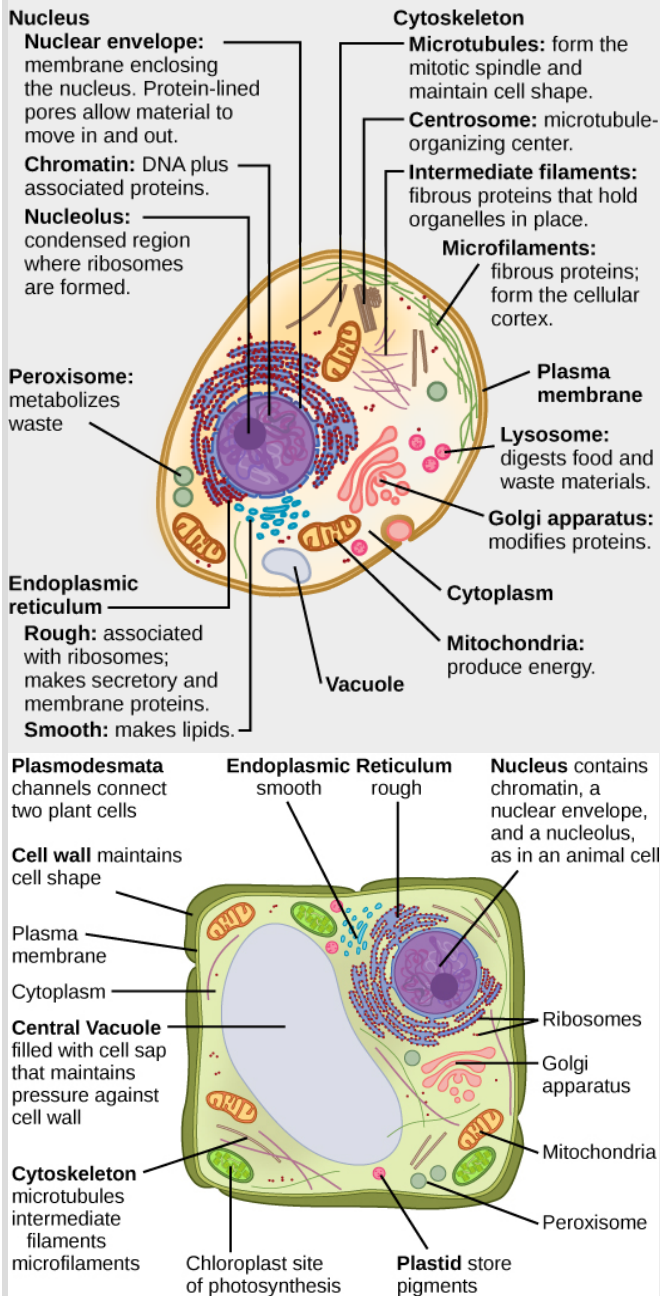
Our natural world also utilizes the principle of form following function, especially in cell biology, and this will become clear as we explore eukaryotic cells ([link](#)). Unlike prokaryotic cells, **eukaryotic cells** have: 1) a membrane-bound nucleus; 2) numerous membrane-bound **organelles** such as the endoplasmic reticulum, Golgi apparatus, chloroplasts, mitochondria, and others; and 3) several, rod-shaped chromosomes. Because a eukaryotic cell’s nucleus is surrounded by a membrane, it is often said to have a “true nucleus.” The word “organelle” means “little organ,” and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.

At this point, it should be clear to you that eukaryotic cells have a more complex structure than prokaryotic cells. Organelles allow different functions to be compartmentalized in different areas of the cell. Before turning to organelles, let’s first examine two important components of the cell: the plasma membrane and the cytoplasm.

Note:

Art Connection

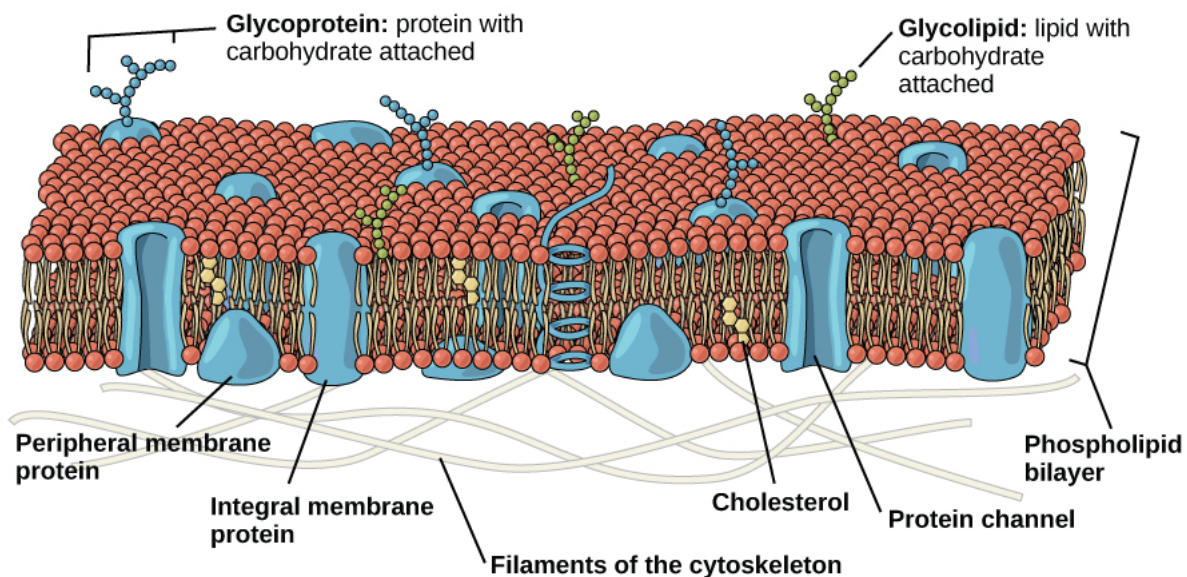
These figures show the major organelles and other cell components of (a) a typical animal cell and (b) a typical eukaryotic plant cell. The plant cell has a cell wall, chloroplasts, plastids, and a central vacuole —structures not found in animal cells. Plant cells do not have lysosomes or centrosomes.



If the nucleolus were not able to carry out its function, what other cellular organelles would be affected?

The Plasma Membrane

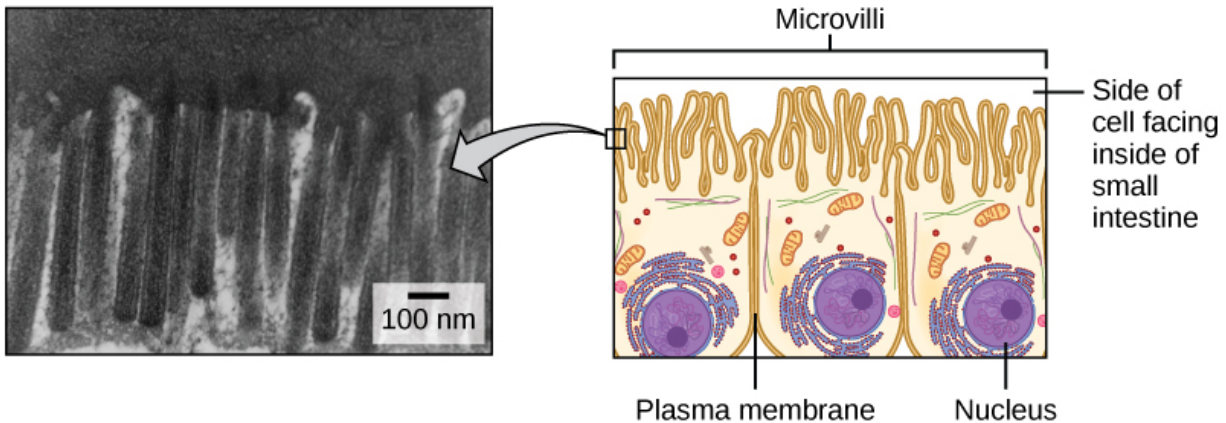
Like prokaryotes, eukaryotic cells have a **plasma membrane** ([\[link\]](#)), a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule with two fatty acid chains and a phosphate-containing group. The plasma membrane controls the passage of organic molecules, ions, water, and oxygen into and out of the cell. Wastes (such as carbon dioxide and ammonia) also leave the cell by passing through the plasma membrane.



The eukaryotic plasma membrane is a phospholipid bilayer with proteins and cholesterol embedded in it.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus); ([\[link\]](#)). Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form following function. People with celiac disease have an immune response to gluten, which is a protein found in wheat, barley, and rye. The immune response damages microvilli, and thus, afflicted individuals cannot absorb

nutrients. This leads to malnutrition, cramping, and diarrhea. Patients suffering from celiac disease must follow a gluten-free diet.



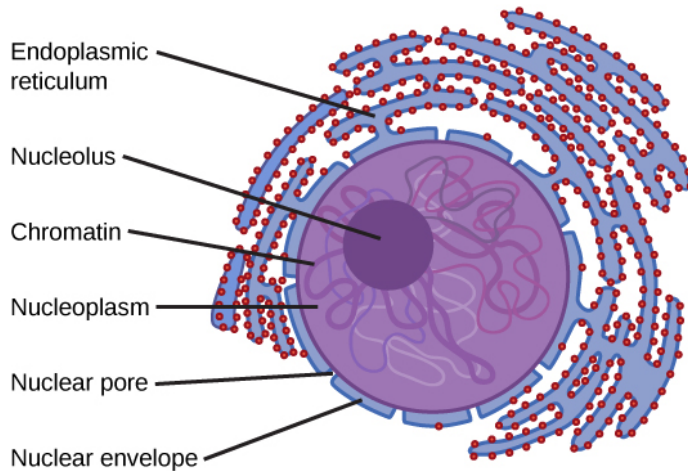
Microvilli, shown here as they appear on cells lining the small intestine, increase the surface area available for absorption. These microvilli are only found on the area of the plasma membrane that faces the cavity from which substances will be absorbed. (credit "micrograph": modification of work by Louisa Howard)

The Cytoplasm

The **cytoplasm** is the entire region of a cell between the plasma membrane and the nuclear envelope (a structure to be discussed shortly). It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals ([\[link\]](#)). Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. However, proteins are not the only organic molecules found in the cytoplasm. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and derivatives of glycerol are found there, too. Ions of sodium, potassium, calcium, and many other elements are also dissolved in the cytoplasm. Many metabolic reactions, including protein synthesis, take place in the cytoplasm.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell ([\[link\]](#)). The **nucleus** (plural = nuclei) houses the cell's DNA and directs the synthesis of ribosomes and proteins. Let's look at it in more detail ([\[link\]](#)).



The nucleus stores chromatin (DNA plus proteins) in a gel-like substance called the nucleoplasm. The nucleolus is a condensed region of chromatin where ribosome synthesis occurs. The boundary of the nucleus is called the nuclear envelope. It consists of two phospholipid bilayers: an outer membrane and an inner membrane. The nuclear membrane is continuous with the endoplasmic reticulum. Nuclear pores allow substances to enter and exit the nucleus.

The Nuclear Envelope

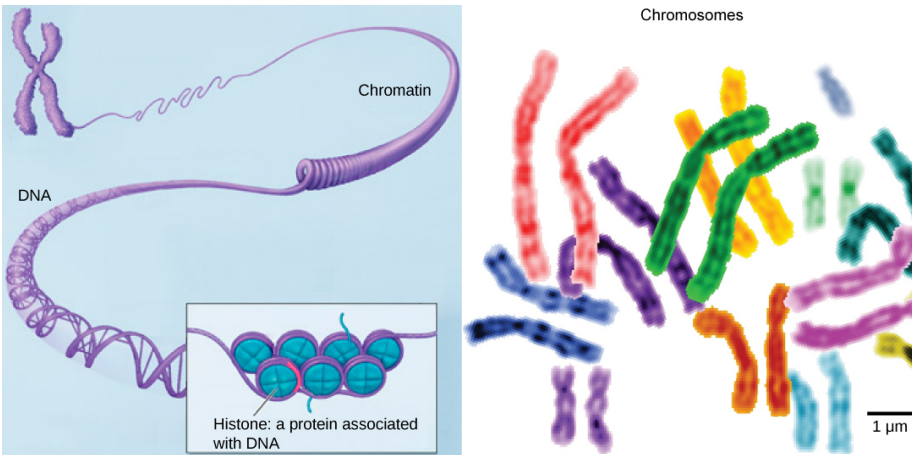
The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus ([\[link\]](#)). Both the inner and outer membranes of the nuclear envelope are phospholipid bilayers.

The nuclear envelope is punctuated with pores that control the passage of ions, molecules, and RNA between the nucleoplasm and cytoplasm. The **nucleoplasm** is the semi-solid fluid inside the nucleus, where we find the chromatin and the nucleolus.

Chromatin and Chromosomes

To understand chromatin, it is helpful to first consider chromosomes.

Chromosomes are structures within the nucleus that are made up of DNA, the hereditary material. You may remember that in prokaryotes, DNA is organized into a single circular chromosome. In eukaryotes, chromosomes are linear structures. Every eukaryotic species has a specific number of chromosomes in the nuclei of its body's cells. For example, in humans, the chromosome number is 46, while in fruit flies, it is eight. Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. When the cell is in the growth and maintenance phases of its life cycle, proteins are attached to chromosomes, and they resemble an unwound, jumbled bunch of threads. These unwound protein-chromosome complexes are called **chromatin** ([\[link\]](#)); chromatin describes the material that makes up the chromosomes both when condensed and decondensed.



(a) This image shows various levels of the organization of chromatin (DNA and protein). (b) This image shows paired chromosomes. (credit b: modification of work by NIH; scale-bar data from Matt Russell)

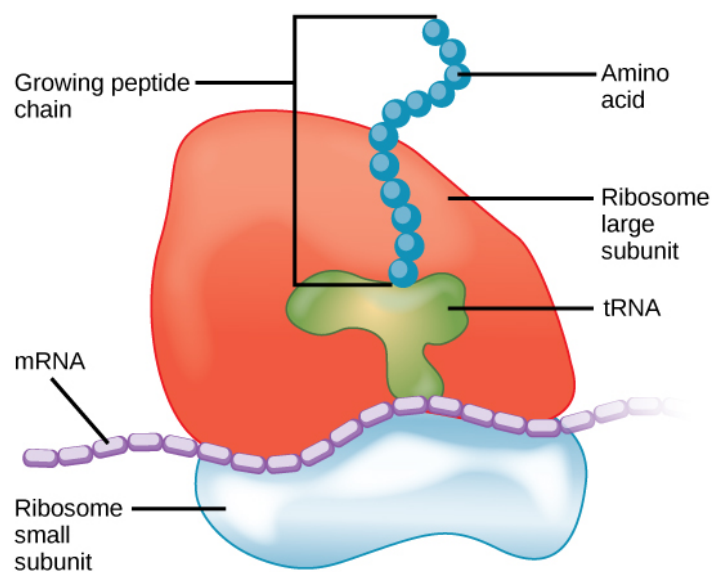
The Nucleolus

We already know that the nucleus directs the synthesis of ribosomes, but how does it do this? Some chromosomes have sections of DNA that encode ribosomal RNA. A darkly staining area within the nucleus called the **nucleolus** (plural = nucleoli) aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported out through the pores in the nuclear envelope to the cytoplasm.

Ribosomes

Ribosomes are the cellular structures responsible for protein synthesis. When viewed through an electron microscope, ribosomes appear either as clusters (polyribosomes) or single, tiny dots that float freely in the cytoplasm. They may be attached to the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum and the

outer membrane of the nuclear envelope ([\[link\]](#)). Electron microscopy has shown us that ribosomes, which are large complexes of protein and RNA, consist of two subunits, aptly called large and small ([\[link\]](#)). Ribosomes receive their “orders” for protein synthesis from the nucleus where the DNA is transcribed into messenger RNA (mRNA). The mRNA travels to the ribosomes, which translate the code provided by the sequence of the nitrogenous bases in the mRNA into a specific order of amino acids in a protein. Amino acids are the building blocks of proteins.



Ribosomes are made up of a large subunit (top) and a small subunit (bottom). During protein synthesis, ribosomes assemble amino acids into proteins.

Because proteins synthesis is an essential function of all cells (including enzymes, hormones, antibodies, pigments, structural components, and surface receptors), ribosomes are found in practically every cell. Ribosomes are particularly abundant in cells that synthesize large amounts of protein. For example, the pancreas is responsible for creating several digestive

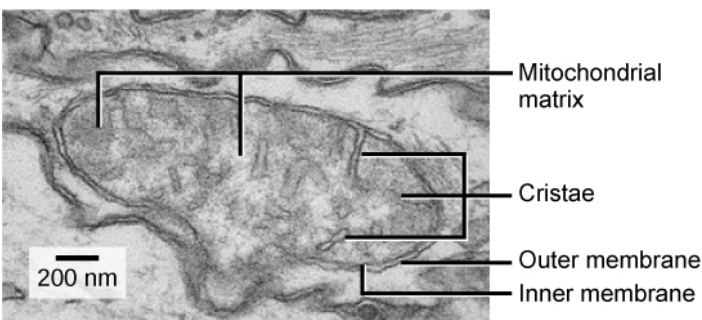
enzymes and the cells that produce these enzymes contain many ribosomes. Thus, we see another example of form following function.

Mitochondria

Mitochondria (singular = mitochondrion) are often called the “powerhouses” or “energy factories” of a cell because they are responsible for making adenosine triphosphate (ATP), the cell’s main energy-carrying molecule. ATP represents the short-term stored energy of the cell. Cellular respiration is the process of making ATP using the chemical energy found in glucose and other nutrients. In mitochondria, this process uses oxygen and produces carbon dioxide as a waste product. In fact, the carbon dioxide that you exhale with every breath comes from the cellular reactions that produce carbon dioxide as a byproduct.

In keeping with our theme of form following function, it is important to point out that muscle cells have a very high concentration of mitochondria that produce ATP. Your muscle cells need a lot of energy to keep your body moving. When your cells don’t get enough oxygen, they do not make a lot of ATP. Instead, the small amount of ATP they make in the absence of oxygen is accompanied by the production of lactic acid.

Mitochondria are oval-shaped, double membrane organelles ([\[link\]](#)) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner layer has folds called cristae. The area surrounded by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.



This electron micrograph shows a mitochondrion as viewed with a transmission electron microscope. This organelle has an outer membrane and an inner membrane. The inner membrane contains folds, called cristae, which increase its surface area. The space between the two membranes is called the intermembrane space, and the space inside the inner membrane is called the mitochondrial matrix. ATP synthesis takes place on the inner membrane. (credit: modification of work by Matthew Britton; scale-bar data from Matt Russell)

Peroxisomes

Peroxisomes are small, round organelles enclosed by single membranes. They carry out oxidation reactions that break down fatty acids and amino acids. They also detoxify many poisons that may enter the body. (Many of these oxidation reactions release hydrogen peroxide, H_2O_2 , which would be damaging to cells; however, when these reactions are confined to peroxisomes, enzymes safely break down the H_2O_2 into oxygen and water.) For example, alcohol is detoxified by peroxisomes in liver cells. Glyoxysomes, which are specialized peroxisomes in plants, are responsible for converting stored fats into sugars.

Vesicles and Vacuoles

Vesicles and **vacuoles** are membrane-bound sacs that function in storage and transport. Other than the fact that vacuoles are somewhat larger than vesicles, there is a very subtle distinction between them: The membranes of

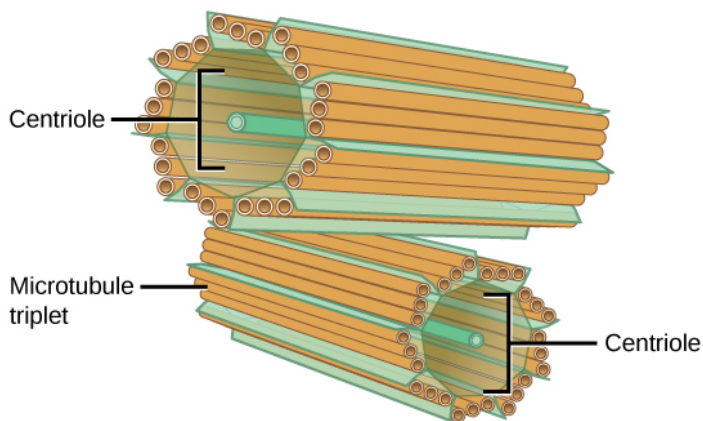
vesicles can fuse with either the plasma membrane or other membrane systems within the cell. Additionally, some agents such as enzymes within plant vacuoles break down macromolecules. The membrane of a vacuole does not fuse with the membranes of other cellular components.

Animal Cells versus Plant Cells

At this point, you know that each eukaryotic cell has a plasma membrane, cytoplasm, a nucleus, ribosomes, mitochondria, peroxisomes, and in some, vacuoles, but there are some striking differences between animal and plant cells. While both animal and plant cells have microtubule organizing centers (MTOCs), animal cells also have centrioles associated with the MTOC: a complex called the centrosome. Animal cells each have a centrosome and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts and other specialized plastids, and a large central vacuole, whereas animal cells do not.

The Centrosome

The **centrosome** is a microtubule-organizing center found near the nuclei of animal cells. It contains a pair of centrioles, two structures that lie perpendicular to each other ([\[link\]](#)). Each centriole is a cylinder of nine triplets of microtubules.



The centrosome consists of two centrioles that lie at right angles to each other. Each centriole is a cylinder made up of nine triplets of microtubules. Nontubulin proteins (indicated by the green lines) hold the microtubule triplets together.

The centrosome (the organelle where all microtubules originate) replicates itself before a cell divides, and the centrioles appear to have some role in pulling the duplicated chromosomes to opposite ends of the dividing cell. However, the exact function of the centrioles in cell division isn't clear, because cells that have had the centrosome removed can still divide, and plant cells, which lack centrosomes, are capable of cell division.

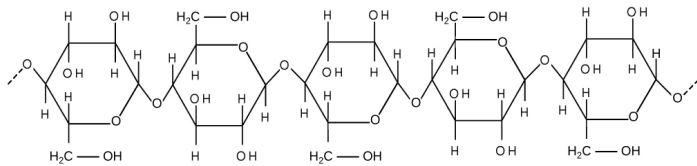
Lysosomes

Animal cells have another set of organelles not found in plant cells: lysosomes. The **lysosomes** are the cell's "garbage disposal." In plant cells, the digestive processes take place in vacuoles. Enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. These enzymes are active at a much lower pH than that of the cytoplasm. Therefore, the pH within lysosomes is more acidic than the pH of the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, so again, the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

The Cell Wall

If you examine [\[link\]](#)**b**, the diagram of a plant cell, you will see a structure external to the plasma membrane called the cell wall. The **cell wall** is a rigid covering that protects the cell, provides structural support, and gives

shape to the cell. Fungal and protistan cells also have cell walls. While the chief component of prokaryotic cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose ([link](#)), a polysaccharide made up of glucose units. Have you ever noticed that when you bite into a raw vegetable, like celery, it crunches? That's because you are tearing the rigid cell walls of the celery cells with your teeth.



Cellulose is a long chain of β -glucose molecules connected by a 1-4 linkage.

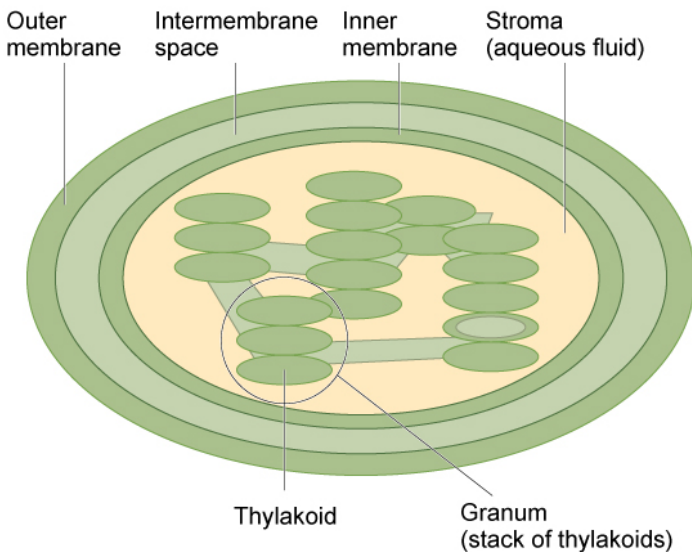
The dashed lines at each end of the figure indicate a series of many more glucose units. The size of the page makes it impossible to portray an entire cellulose molecule.

Chloroplasts

Like the mitochondria, chloroplasts have their own DNA and ribosomes, but chloroplasts have an entirely different function. **Chloroplasts** are plant cell organelles that carry out photosynthesis. Photosynthesis is the series of reactions that use carbon dioxide, water, and light energy to make glucose and oxygen. This is a major difference between plants and animals; plants (autotrophs) are able to make their own food, like sugars, while animals (heterotrophs) must ingest their food.

Like mitochondria, chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of

interconnected and stacked fluid-filled membrane sacs called thylakoids ([\[link\]](#)). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane that surrounds the grana is called the stroma.



The chloroplast has an outer membrane, an inner membrane, and membrane structures called thylakoids that are stacked into grana. The space inside the thylakoid membranes is called the thylakoid space. The light harvesting reactions take place in the thylakoid membranes, and the synthesis of sugar takes place in the fluid inside the inner membrane, which is called the stroma. Chloroplasts also have their own genome, which is contained on a single circular chromosome.

The chloroplasts contain a green pigment called **chlorophyll**, which captures the light energy that drives the reactions of photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria perform photosynthesis, but their chlorophyll is not relegated to an organelle.

Note:

Evolution Connection

Endosymbiosis

We have mentioned that both mitochondria and chloroplasts contain DNA and ribosomes. Have you wondered why? Strong evidence points to endosymbiosis as the explanation.

Symbiosis is a relationship in which organisms from two separate species depend on each other for their survival. Endosymbiosis (endo- = “within”) is a mutually beneficial relationship in which one organism lives inside the other. Endosymbiotic relationships abound in nature. We have already mentioned that microbes that produce vitamin K live inside the human gut. This relationship is beneficial for us because we are unable to synthesize vitamin K. It is also beneficial for the microbes because they are protected from other organisms and from drying out, and they receive abundant food from the environment of the large intestine.

Scientists have long noticed that bacteria, mitochondria, and chloroplasts are similar in size. We also know that bacteria have DNA and ribosomes, just as mitochondria and chloroplasts do. Scientists believe that host cells and bacteria formed an endosymbiotic relationship when the host cells ingested both aerobic and autotrophic bacteria (cyanobacteria) but did not destroy them. Through many millions of years of evolution, these ingested bacteria became more specialized in their functions, with the aerobic bacteria becoming mitochondria and the autotrophic bacteria becoming chloroplasts.

The Central Vacuole

Previously, we mentioned vacuoles as essential components of plant cells. If you look at [link](#)b, you will see that plant cells each have a large central vacuole that occupies most of the area of the cell. The **central vacuole** plays a key role in regulating the cell's concentration of water in changing environmental conditions. Have you ever noticed that if you forget to water a plant for a few days, it wilts? That's because as the water concentration in the soil becomes lower than the water concentration in the plant, water moves out of the central vacuoles and cytoplasm. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of plant cells results in the wilted appearance of the plant.

The central vacuole also supports the expansion of the cell. When the central vacuole holds more water, the cell gets larger without having to invest a lot of energy in synthesizing new cytoplasm.

Section Summary

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. The plasma membrane is a phospholipid bilayer embedded with proteins. The nucleus's nucleolus is the site of ribosome assembly. Ribosomes are either found in the cytoplasm or attached to the cytoplasmic side of the plasma membrane or endoplasmic reticulum. They perform protein synthesis. Mitochondria participate in cellular respiration; they are responsible for the majority of ATP produced in the cell. Peroxisomes hydrolyze fatty acids, amino acids, and some toxins. Vesicles and vacuoles are storage and transport compartments. In plant cells, vacuoles also help break down macromolecules.

Animal cells also have a centrosome and lysosomes. The centrosome has two bodies perpendicular to each other, the centrioles, and has an unknown purpose in cell division. Lysosomes are the digestive organelles of animal cells.

Plant cells and plant-like cells each have a cell wall, chloroplasts, and a central vacuole. The plant cell wall, whose primary component is cellulose, protects the cell, provides structural support, and gives shape to the cell. Photosynthesis takes place in chloroplasts. The central vacuole can expand without having to produce more cytoplasm.

Art Connections

Exercise:

Problem:

[\[link\]](#) If the nucleolus were not able to carry out its function, what other cellular organelles would be affected?

Solution:

[\[link\]](#) Free ribosomes and rough endoplasmic reticulum (which contains ribosomes) would not be able to form.

Review Questions

Exercise:

Problem:

Which of the following is surrounded by two phospholipid bilayers?

- a. the ribosomes
- b. the vesicles
- c. the cytoplasm
- d. the nucleoplasm

Solution:

D

Exercise:

Problem: Peroxisomes got their name because hydrogen peroxide is:

- a. used in their detoxification reactions
- b. produced during their oxidation reactions
- c. incorporated into their membranes
- d. a cofactor for the organelles' enzymes

Solution:

B

Exercise:

Problem:

In plant cells, the function of the lysosomes is carried out by _____.

- a. vacuoles
- b. peroxisomes
- c. ribosomes
- d. nuclei

Solution:

A

Exercise:

Problem:

Which of the following is found both in eukaryotic and prokaryotic cells?

- a. nucleus
- b. mitochondrion

- c. vacuole
- d. ribosomes

Solution:

D

Free Response

Exercise:

Problem:

You already know that ribosomes are abundant in red blood cells. In what other cells of the body would you find them in great abundance? Why?

Solution:

Ribosomes are abundant in muscle cells as well because muscle cells are constructed of the proteins made by the ribosomes.

Exercise:

Problem:

What are the structural and functional similarities and differences between mitochondria and chloroplasts?

Solution:

Both are similar in that they are enveloped in a double membrane, both have an intermembrane space, and both make ATP. Both mitochondria and chloroplasts have DNA, and mitochondria have inner folds called cristae and a matrix, while chloroplasts have chlorophyll and accessory pigments in the thylakoids that form stacks (grana) and a stroma.

Glossary

cell wall

rigid cell covering made of various molecules that protects the cell, provides structural support, and gives shape to the cell

central vacuole

large plant cell organelle that regulates the cell's storage compartment, holds water, and plays a significant role in cell growth as the site of macromolecule degradation

centrosome

region in animal cells made of two centrioles

chlorophyll

green pigment that captures the light energy that drives the light reactions of photosynthesis

chloroplast

plant cell organelle that carries out photosynthesis

chromatin

protein-DNA complex that serves as the building material of chromosomes

chromosome

structure within the nucleus that is made up of chromatin that contains DNA, the hereditary material

cytoplasm

entire region between the plasma membrane and the nuclear envelope, consisting of organelles suspended in the gel-like cytosol, the cytoskeleton, and various chemicals

cytosol

gel-like material of the cytoplasm in which cell structures are suspended

eukaryotic cell

cell that has a membrane-bound nucleus and several other membrane-bound compartments or sacs

lysosome

organelle in an animal cell that functions as the cell's digestive component; it breaks down proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles

mitochondria

(singular = mitochondrion) cellular organelles responsible for carrying out cellular respiration, resulting in the production of ATP, the cell's main energy-carrying molecule

nuclear envelope

double-membrane structure that constitutes the outermost portion of the nucleus

nucleolus

darkly staining body within the nucleus that is responsible for assembling the subunits of the ribosomes

nucleoplasm

semi-solid fluid inside the nucleus that contains the chromatin and nucleolus

nucleus

cell organelle that houses the cell's DNA and directs the synthesis of ribosomes and proteins

organelle

compartment or sac within a cell

peroxisome

small, round organelle that contains hydrogen peroxide, oxidizes fatty acids and amino acids, and detoxifies many poisons

plasma membrane

phospholipid bilayer with embedded (integral) or attached (peripheral) proteins, and separates the internal content of the cell from its surrounding environment

ribosome

cellular structure that carries out protein synthesis

vacuole

membrane-bound sac, somewhat larger than a vesicle, which functions in cellular storage and transport

vesicle

small, membrane-bound sac that functions in cellular storage and transport; its membrane is capable of fusing with the plasma membrane and the membranes of the endoplasmic reticulum and Golgi apparatus

The Endomembrane System and Proteins

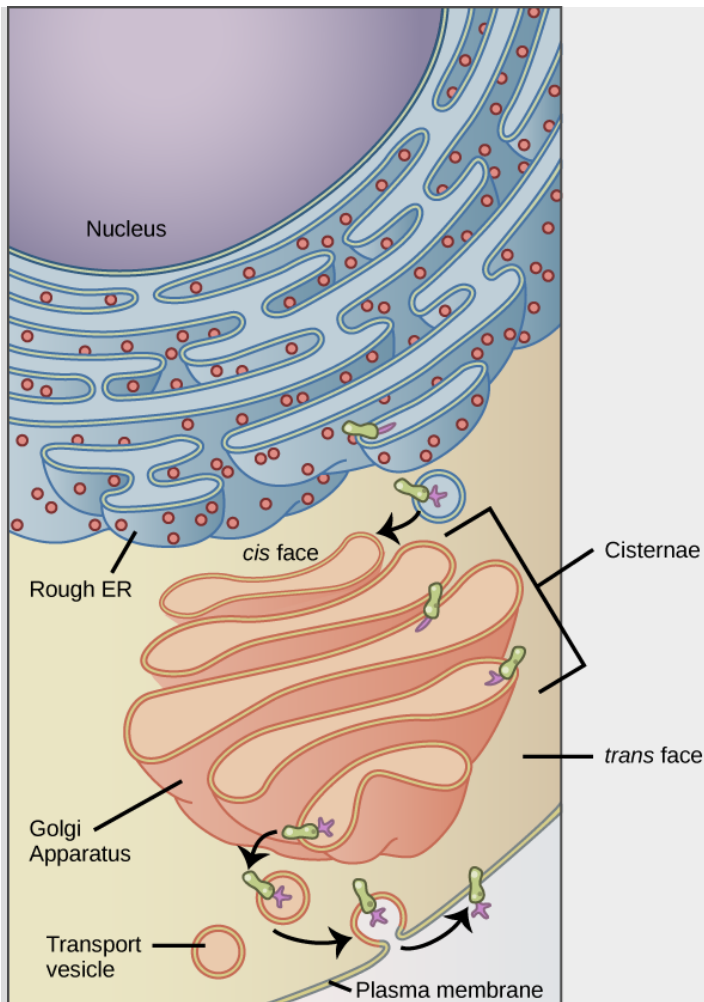
By the end of this section, you will be able to:

- List the components of the endomembrane system
- Recognize the relationship between the endomembrane system and its functions

The endomembrane system (endo = “within”) is a group of membranes and organelles ([\[link\]](#)) in eukaryotic cells that works together to modify, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, which we’ve already mentioned, and the endoplasmic reticulum and Golgi apparatus, which we will cover shortly. Although not technically *within* the cell, the plasma membrane is included in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles. The endomembrane system does not include the membranes of either mitochondria or chloroplasts.

Note:

Art Connection



"Membrane and secretory proteins are synthesized in the rough endoplasmic reticulum (RER). The RER also sometimes modifies proteins. In this illustration, a (green) integral membrane protein in the ER is modified by attachment of a (purple) carbohydrate. Vesicles with the integral protein bud from the ER and fuse with the cis face of the Golgi apparatus. As the protein passes along the Golgi's cisternae, it is further modified by the addition of more carbohydrates. After its synthesis is

complete, it exits as integral membrane protein of the vesicle that bud from the Golgi's **trans** face and when the vesicle fuses with the cell membrane the protein becomes integral portion of that cell membrane.
(credit: modification of work by Magnus Manske)

If a peripheral membrane protein were synthesized in the lumen (inside) of the ER, would it end up on the inside or outside of the plasma membrane?

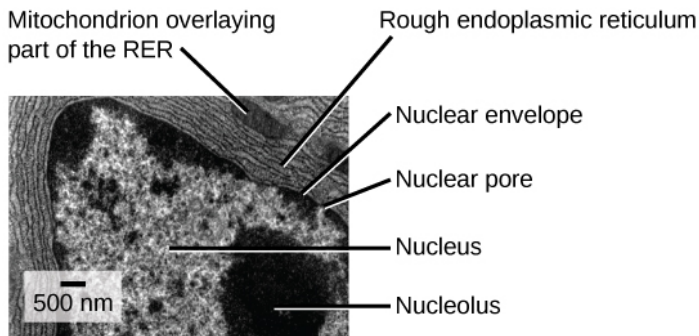
The Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** ([\[link\]](#)) is a series of interconnected membranous sacs and tubules that collectively modifies proteins and synthesizes lipids. However, these two functions are performed in separate areas of the ER: the rough ER and the smooth ER, respectively.

The hollow portion of the ER tubules is called the lumen or cisternal space. The membrane of the ER, which is a phospholipid bilayer embedded with proteins, is continuous with the nuclear envelope.

Rough ER

The **rough endoplasmic reticulum (RER)** is so named because the ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope ([\[link\]](#)).



This transmission electron micrograph shows the rough endoplasmic reticulum and other organelles in a pancreatic cell. (credit: modification of work by Louisa Howard)

Ribosomes transfer their newly synthesized proteins into the lumen of the RER where they undergo structural modifications, such as folding or the acquisition of side chains. These modified proteins will be incorporated into cellular membranes—the membrane of the ER or those of other organelles—or secreted from the cell (such as protein hormones, enzymes). The RER also makes phospholipids for cellular membranes.

If the phospholipids or modified proteins are not destined to stay in the RER, they will reach their destinations via transport vesicles that bud from the RER's membrane ([\[link\]](#)).

Since the RER is engaged in modifying proteins (such as enzymes, for example) that will be secreted from the cell, you would be correct in assuming that the RER is abundant in cells that secrete proteins. This is the case with cells of the liver, for example.

Smooth ER

The **smooth endoplasmic reticulum (SER)** is continuous with the RER but has few or no ribosomes on its cytoplasmic surface ([\[link\]](#)). Functions

of the SER include synthesis of carbohydrates, lipids, and steroid hormones; detoxification of medications and poisons; and storage of calcium ions.

In muscle cells, a specialized SER called the sarcoplasmic reticulum is responsible for storage of the calcium ions that are needed to trigger the coordinated contractions of the muscle cells.

Note:

Link to Learning



You can watch an excellent animation of the endomembrane system [here](#). At the end of the animation, there is a short self-assessment.

Note:

Career Connection

Cardiologist

Heart disease is the leading cause of death in the United States. This is primarily due to our sedentary lifestyle and our high trans-fat diets.

Heart failure is just one of many disabling heart conditions. Heart failure does not mean that the heart has stopped working. Rather, it means that the heart can't pump with sufficient force to transport oxygenated blood to all the vital organs. Left untreated, heart failure can lead to kidney failure and failure of other organs.

The wall of the heart is composed of cardiac muscle tissue. Heart failure occurs when the endoplasmic reticula of cardiac muscle cells do not

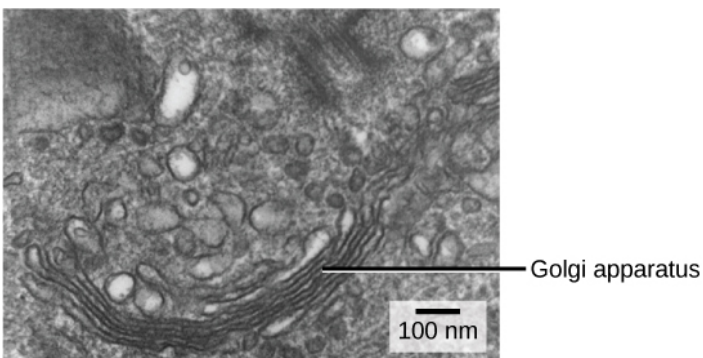
function properly. As a result, an insufficient number of calcium ions are available to trigger a sufficient contractile force.

Cardiologists (cardi- = “heart”; -ologist = “one who studies”) are doctors who specialize in treating heart diseases, including heart failure.

Cardiologists can make a diagnosis of heart failure via physical examination, results from an electrocardiogram (ECG, a test that measures the electrical activity of the heart), a chest X-ray to see whether the heart is enlarged, and other tests. If heart failure is diagnosed, the cardiologist will typically prescribe appropriate medications and recommend a reduction in table salt intake and a supervised exercise program.

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER and transport their contents elsewhere, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles still need to be sorted, packaged, and tagged so that they wind up in the right place. Sorting, tagging, packaging, and distribution of lipids and proteins takes place in the **Golgi apparatus** (also called the Golgi body), a series of flattened membranes ([\[link\]](#)).



The Golgi apparatus in this white blood cell is visible as a stack of semicircular, flattened rings in the lower portion of the image. Several

vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard)

The receiving side of the Golgi apparatus is called the *cis* face. The opposite side is called the *trans* face. The transport vesicles that formed from the ER travel to the *cis* face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further modifications that allow them to be sorted. The most frequent modification is the addition of short chains of sugar molecules. These newly modified proteins and lipids are then tagged with phosphate groups or other small molecules so that they can be routed to their proper destinations.

Finally, the modified and tagged proteins are packaged into secretory vesicles that bud from the *trans* face of the Golgi. While some of these vesicles deposit their contents into other parts of the cell where they will be used, other secretory vesicles fuse with the plasma membrane and release their contents outside the cell.

In another example of form following function, cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundance of Golgi.

In plant cells, the Golgi apparatus has the additional role of synthesizing polysaccharides, some of which are incorporated into the cell wall and some of which are used in other parts of the cell.

Note:**Career Connection****Geneticist**

Many diseases arise from genetic mutations that prevent the synthesis of critical proteins. One such disease is Lowe disease (also called

oculocerebrorenal syndrome, because it affects the eyes, brain, and kidneys). In Lowe disease, there is a deficiency in an enzyme localized to the Golgi apparatus. Children with Lowe disease are born with cataracts, typically develop kidney disease after the first year of life, and may have impaired mental abilities.

Lowe disease is a genetic disease caused by a mutation on the X chromosome. The X chromosome is one of the two human sex chromosomes, as these chromosomes determine a person's sex. Females possess two X chromosomes while males possess one X and one Y chromosome. In females, the genes on only one of the two X chromosomes are expressed. Therefore, females who carry the Lowe disease gene on one of their X chromosomes have a 50/50 chance of having the disease.

However, males only have one X chromosome and the genes on this chromosome are always expressed. Therefore, males will always have Lowe disease if their X chromosome carries the Lowe disease gene. The location of the mutated gene, as well as the locations of many other mutations that cause genetic diseases, has now been identified. Through prenatal testing, a woman can find out if the fetus she is carrying may be afflicted with one of several genetic diseases.

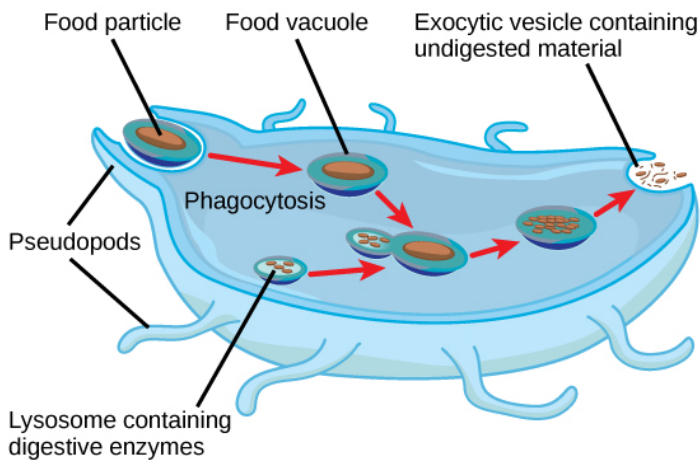
Geneticists analyze the results of prenatal genetic tests and may counsel pregnant women on available options. They may also conduct genetic research that leads to new drugs or foods, or perform DNA analyses that are used in forensic investigations.

Lysosomes

In addition to their role as the digestive component and organelle-recycling facility of animal cells, lysosomes are considered to be parts of the endomembrane system. Lysosomes also use their hydrolytic enzymes to destroy pathogens (disease-causing organisms) that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis or endocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from

the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen ([link](#)).

Phagocytosis



A macrophage has engulfed (phagocytized) a potentially pathogenic bacterium and then fuses with a lysosomes within the cell to destroy the pathogen. Other organelles are present in the cell but for simplicity are not shown.

Section Summary

The endomembrane system includes the nuclear envelope, lysosomes, vesicles, the ER, and Golgi apparatus, as well as the plasma membrane. These cellular components work together to modify, package, tag, and transport proteins and lipids that form the membranes.

The RER modifies proteins and synthesizes phospholipids used in cell membranes. The SER synthesizes carbohydrates, lipids, and steroid

hormones; engages in the detoxification of medications and poisons; and stores calcium ions. Sorting, tagging, packaging, and distribution of lipids and proteins take place in the Golgi apparatus. Lysosomes are created by the budding of the membranes of the RER and Golgi. Lysosomes digest macromolecules, recycle worn-out organelles, and destroy pathogens.

Art Connections

Exercise:

Problem:

[\[link\]](#) If a peripheral membrane protein were synthesized in the lumen (inside) of the ER, would it end up on the inside or outside of the plasma membrane?

Solution:

[\[link\]](#) It would end up on the outside. After the vesicle passes through the Golgi apparatus and fuses with the plasma membrane, it turns inside out.

Review Questions

Exercise:

Problem:

Which of the following is not a component of the endomembrane system?

- a. mitochondrion
 - b. Golgi apparatus
 - c. endoplasmic reticulum
 - d. lysosome
-

Solution:

A

Exercise:

Problem:

The process by which a cell engulfs a foreign particle is known as:

- a. endosymbiosis
- b. phagocytosis
- c. hydrolysis
- d. membrane synthesis

Solution:

B

Exercise:

Problem:

Which of the following is most likely to have the greatest concentration of smooth endoplasmic reticulum?

- a. a cell that secretes enzymes
- b. a cell that destroys pathogens
- c. a cell that makes steroid hormones
- d. a cell that engages in photosynthesis

Solution:

C

Exercise:

Problem:

Which of the following sequences correctly lists in order the steps involved in the incorporation of a proteinaceous molecule within a cell?

- a. synthesis of the protein on the ribosome; modification in the Golgi apparatus; packaging in the endoplasmic reticulum; tagging in the vesicle
 - b. synthesis of the protein on the lysosome; tagging in the Golgi; packaging in the vesicle; distribution in the endoplasmic reticulum
 - c. synthesis of the protein on the ribosome; modification in the endoplasmic reticulum; tagging in the Golgi; distribution via the vesicle
 - d. synthesis of the protein on the lysosome; packaging in the vesicle; distribution via the Golgi; tagging in the endoplasmic reticulum
-

Solution:

C

Free Response

Exercise:

Problem:

In the context of cell biology, what do we mean by form follows function? What are at least two examples of this concept?

Solution:

“Form follows function” refers to the idea that the function of a body part dictates the form of that body part. As an example, compare your arm to a bat’s wing. While the bones of the two correspond, the parts serve different functions in each organism and their forms have adapted to follow that function.

Exercise:

Problem:

In your opinion, is the nuclear membrane part of the endomembrane system? Why or why not? Defend your answer.

Solution:

Since the external surface of the nuclear membrane is continuous with the rough endoplasmic reticulum, which is part of the endomembrane system, then it is correct to say that it is part of the system.

Glossary

endomembrane system

group of organelles and membranes in eukaryotic cells that work together modifying, packaging, and transporting lipids and proteins

endoplasmic reticulum (ER)

series of interconnected membranous structures within eukaryotic cells that collectively modify proteins and synthesize lipids

Golgi apparatus

eukaryotic organelle made up of a series of stacked membranes that sorts, tags, and packages lipids and proteins for distribution

rough endoplasmic reticulum (RER)

region of the endoplasmic reticulum that is studded with ribosomes and engages in protein modification and phospholipid synthesis

smooth endoplasmic reticulum (SER)

region of the endoplasmic reticulum that has few or no ribosomes on its cytoplasmic surface and synthesizes carbohydrates, lipids, and steroid hormones; detoxifies certain chemicals (like pesticides, preservatives, medications, and environmental pollutants), and stores calcium ions

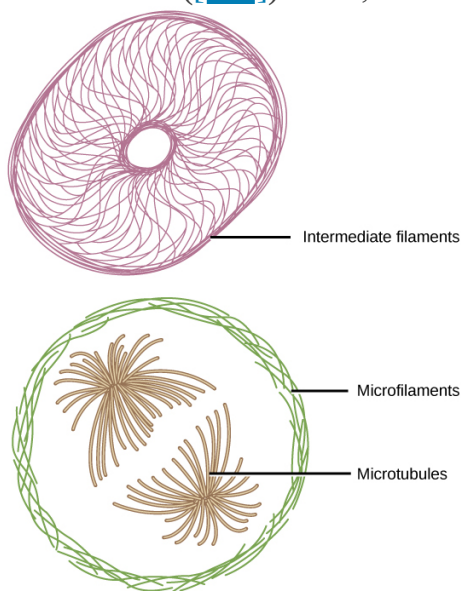
The Cytoskeleton

By the end of this section, you will be able to:

- Describe the cytoskeleton
- Compare the roles of microfilaments, intermediate filaments, and microtubules
- Compare and contrast cilia and flagella
- Summarize the differences among the components of prokaryotic cells, animal cells, and plant cells

If you were to remove all the organelles from a cell, would the plasma membrane and the cytoplasm be the only components left? No. Within the cytoplasm, there would still be ions and organic molecules, plus a network of protein fibers that help maintain the shape of the cell, secure some organelles in specific positions, allow cytoplasm and vesicles to move within the cell, and enable cells within multicellular organisms to move.

Collectively, this network of protein fibers is known as the **cytoskeleton**. There are three types of fibers within the cytoskeleton: microfilaments, intermediate filaments, and microtubules ([\[link\]](#)). Here, we will examine each.



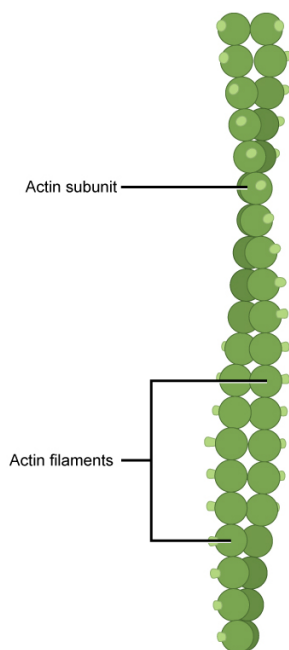
Microfilaments thicken the cortex around the inner edge of a cell; like rubber bands, they resist tension.

Microtubules are found in the interior of the cell where they maintain cell shape by resisting compressive forces.

Intermediate filaments are found throughout the cell and hold organelles in place.

Microfilaments

Of the three types of protein fibers in the cytoskeleton, **microfilaments** are the narrowest. They function in cellular movement, have a diameter of about 7 nm, and are made of two intertwined strands of a globular protein called actin ([link](#)). For this reason, microfilaments are also known as actin filaments.



Microfilaments are made of two intertwined strands of actin.

Actin is powered by ATP to assemble its filamentous form, which serves as a track for the movement of a motor protein called myosin. This enables actin to engage in cellular events requiring motion, such as cell division in animal cells and cytoplasmic streaming, which is the circular movement of the cell cytoplasm in plant cells. Actin and myosin are plentiful in muscle cells. When your actin and myosin filaments slide past each other, your muscles contract.

Microfilaments also provide some rigidity and shape to the cell. They can depolymerize (disassemble) and reform quickly, thus enabling a cell to change its shape and move.

White blood cells (your body's infection-fighting cells) make good use of this ability. They can move to the site of an infection and phagocytize the pathogen.

Note:

Link to Learning



To see an example of a white blood cell in action, watch a short time-lapse video of the cell capturing two bacteria. It engulfs one and then moves on to the other.

https://www.openstaxcollege.org/l/chasing_bacteria

Intermediate Filaments

Intermediate filaments are made of several strands of fibrous proteins that are wound together ([link]). These elements of the cytoskeleton get their name from the fact that their diameter, 8 to 10 nm, is between those of microfilaments and microtubules.



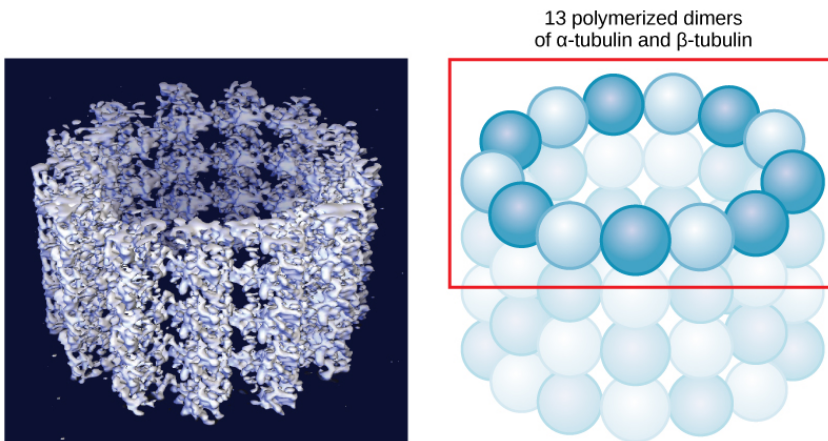
Intermediate filaments consist of several intertwined strands of fibrous proteins.

Intermediate filaments have no role in cell movement. Their function is purely structural. They bear tension, thus maintaining the shape of the cell, and anchor the nucleus and other organelles in place. [link] shows how intermediate filaments create a supportive scaffolding inside the cell.

The intermediate filaments are the most diverse group of cytoskeletal elements. Several types of fibrous proteins are found in the intermediate filaments. You are probably most familiar with keratin, the fibrous protein that strengthens your hair, nails, and the epidermis of the skin.

Microtubules

As their name implies, microtubules are small hollow tubes. The walls of the microtubule are made of polymerized dimers of α -tubulin and β -tubulin, two globular proteins ([\[link\]](#)). With a diameter of about 25 nm, **microtubules** are the widest components of the cytoskeleton. They help the cell resist compression, provide a track along which vesicles move through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. Like microfilaments, microtubules can dissolve and reform quickly.



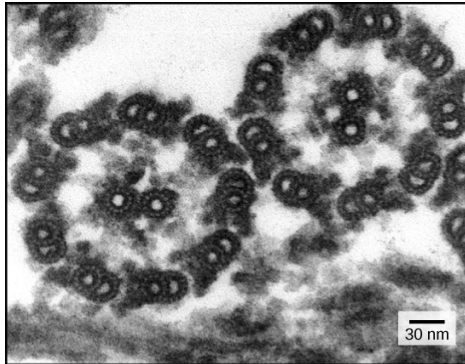
Microtubules are hollow. Their walls consist of 13 polymerized dimers of α -tubulin and β -tubulin (right image). The left image shows the molecular structure of the tube.

Microtubules are also the structural elements of flagella, cilia, and centrioles (the latter are the two perpendicular bodies of the centrosome). In fact, in animal cells, the centrosome is the microtubule-organizing center. In eukaryotic cells, flagella and cilia are quite different structurally from their counterparts in prokaryotes, as discussed below.

Flagella and Cilia

To refresh your memory, **flagella** (singular = flagellum) are long, hair-like structures that extend from the plasma membrane and are used to move an entire cell (for example, sperm, *Euglena*). When present, the cell has just one flagellum or a few flagella. When **cilia** (singular = cilium) are present, however, many of them extend along the entire surface of the plasma membrane. They are short, hair-like structures that are used to move entire cells (such as paramecia) or substances along the outer surface of the cell (for example, the cilia of cells lining the Fallopian tubes that move the ovum toward the uterus, or cilia lining the cells of the respiratory tract that trap particulate matter and move it toward your nostrils.)

Despite their differences in length and number, flagella and cilia share a common structural arrangement of microtubules called a “9 + 2 array.” This is an appropriate name because a single flagellum or cilium is made of a ring of nine microtubule doublets, surrounding a single microtubule doublet in the center ([\[link\]](#)).



This transmission electron micrograph of two flagella shows the 9 + 2 array of microtubules: nine microtubule doublets surround a single microtubule doublet. (credit: modification of work by Dartmouth Electron Microscope Facility, Dartmouth College; scale-bar data from Matt Russell)

You have now completed a broad survey of the components of prokaryotic and eukaryotic cells. For a summary of cellular components in prokaryotic and eukaryotic cells, see [\[link\]](#).

Components of Prokaryotic and Eukaryotic Cells				
Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Plasma membrane	Separates cell from external environment; controls passage of organic molecules, ions, water, oxygen, and wastes into and out of cell	Yes	Yes	Yes
Cytoplasm	Provides turgor pressure to plant cells as fluid inside the central vacuole; site of many metabolic reactions; medium in which organelles are found	Yes	Yes	Yes
Nucleolus	Darkened area within the nucleus where ribosomal subunits are synthesized.	No	Yes	Yes
Nucleus	Cell organelle that houses DNA and directs synthesis of ribosomes and proteins	No	Yes	Yes
Ribosomes	Protein synthesis	Yes	Yes	Yes

Components of Prokaryotic and Eukaryotic Cells				
Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Mitochondria	ATP production/cellular respiration	No	Yes	Yes
Peroxisomes	Oxidizes and thus breaks down fatty acids and amino acids, and detoxifies poisons	No	Yes	Yes
Vesicles and vacuoles	Storage and transport; digestive function in plant cells	No	Yes	Yes
Centrosome	Unspecified role in cell division in animal cells; source of microtubules in animal cells	No	Yes	No
Lysosomes	Digestion of macromolecules; recycling of worn-out organelles	No	Yes	No
Cell wall	Protection, structural support and maintenance of cell shape	Yes, primarily peptidoglycan	No	Yes, primarily cellulose
Chloroplasts	Photosynthesis	No	No	Yes

Components of Prokaryotic and Eukaryotic Cells				
Cell Component	Function	Present in Prokaryotes?	Present in Animal Cells?	Present in Plant Cells?
Endoplasmic reticulum	Modifies proteins and synthesizes lipids	No	Yes	Yes
Golgi apparatus	Modifies, sorts, tags, packages, and distributes lipids and proteins	No	Yes	Yes
Cytoskeleton	Maintains cell's shape, secures organelles in specific positions, allows cytoplasm and vesicles to move within cell, and enables unicellular organisms to move independently	Yes	Yes	Yes
Flagella	Cellular locomotion	Some	Some	No, except for some plant sperm cells.
Cilia	Cellular locomotion, movement of particles along extracellular surface of plasma membrane, and filtration	Some	Some	No

Section Summary

The cytoskeleton has three different types of protein elements. From narrowest to widest, they are the microfilaments (actin filaments), intermediate filaments, and microtubules. Microfilaments are often associated with myosin. They provide rigidity and shape to the cell and facilitate cellular movements. Intermediate filaments bear tension and anchor the nucleus and other organelles in place. Microtubules help the cell resist compression, serve as tracks for motor proteins that move vesicles through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. They are also the structural element of centrioles, flagella, and cilia.

Review Questions

Exercise:

Problem:

Which of the following have the ability to disassemble and reform quickly?

- a. microfilaments and intermediate filaments
- b. microfilaments and microtubules
- c. intermediate filaments and microtubules
- d. only intermediate filaments

Solution:

B

Exercise:

Problem: Which of the following do not play a role in intracellular movement?

- a. microfilaments and intermediate filaments
- b. microfilaments and microtubules
- c. intermediate filaments and microtubules
- d. only intermediate filaments

Solution:

D

Free Response

Exercise:**Problem:**

What are the similarities and differences between the structures of centrioles and flagella?

Solution:

Centrioles and flagella are alike in that they are made up of microtubules. In centrioles, two rings of nine microtubule “triplets” are arranged at right angles to one another. This arrangement does not occur in flagella.

Exercise:

Problem:How do cilia and flagella differ?

Solution:

Cilia and flagella are alike in that they are made up of microtubules. Cilia are short, hair-like structures that exist in large numbers and usually cover the entire surface of the plasma membrane. Flagella, in contrast, are long, hair-like structures; when flagella are present, a cell has just one or two.

Glossary**cilium**

(plural = cilia) short, hair-like structure that extends from the plasma membrane in large numbers and is used to move an entire cell or move substances along the outer surface of the cell

cytoskeleton

network of protein fibers that collectively maintain the shape of the cell, secure some organelles in specific positions, allow cytoplasm and vesicles to move within the cell, and enable unicellular organisms to move independently

flagellum

(plural = flagella) long, hair-like structure that extends from the plasma membrane and is used to move the cell

intermediate filament

cytoskeletal component, composed of several intertwined strands of fibrous protein, that bears tension, supports cell-cell junctions, and anchors cells to extracellular structures

microfilament

narrowest element of the cytoskeleton system; it provides rigidity and shape to the cell and enables cellular movements

microtubule

widest element of the cytoskeleton system; it helps the cell resist compression, provides a track along which vesicles move through the cell, pulls replicated chromosomes to opposite ends of a dividing cell, and is the structural element of centrioles, flagella, and cilia

Connections between Cells and Cellular Activities

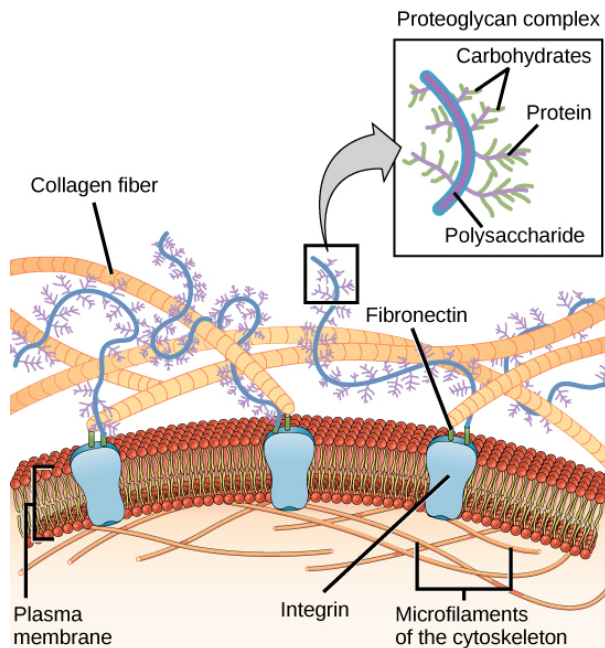
By the end of this section, you will be able to:

- Describe the extracellular matrix
- List examples of the ways that plant cells and animal cells communicate with adjacent cells
- Summarize the roles of tight junctions, desmosomes, gap junctions, and plasmodesmata

You already know that a group of similar cells working together is called a tissue. As you might expect, if cells are to work together, they must communicate with each other, just as you need to communicate with others if you work on a group project. Let's take a look at how cells communicate with each other.

Extracellular Matrix of Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are proteins, and the most abundant protein is collagen. Collagen fibers are interwoven with carbohydrate-containing protein molecules called proteoglycans. Collectively, these materials are called the **extracellular matrix** ([\[link\]](#)). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other. How can this happen?



The extracellular matrix consists of a network of proteins and carbohydrates.

Cells have protein receptors on the extracellular surfaces of their plasma membranes. When a molecule within the matrix binds to the receptor, it changes the molecular structure of the receptor. The receptor, in turn, changes the conformation of the microfilaments positioned just inside the plasma membrane. These conformational changes induce chemical signals inside the cell that reach the nucleus and turn “on” or “off” the transcription of specific sections of DNA, which affects the production of associated proteins, thus changing the activities within the cell.

Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates the adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and

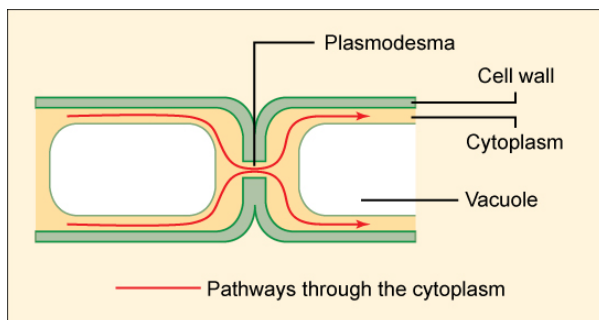
initiates a series of steps that stimulate the platelets to produce clotting factors.

Intercellular Junctions

Cells can also communicate with each other via direct contact, referred to as intercellular junctions. There are some differences in the ways that plant and animal cells do this. Plasmodesmata are junctions between plant cells, whereas animal cell contacts include tight junctions, gap junctions, and desmosomes.

Plasmodesmata

In general, long stretches of the plasma membranes of neighboring plant cells cannot touch one another because they are separated by the cell wall that surrounds each cell ([link](#))^b. How then, can a plant transfer water and other soil nutrients from its roots, through its stems, and to its leaves? Such transport uses the vascular tissues (xylem and phloem) primarily. There also exist structural modifications called **plasmodesmata** (singular = plasmodesma), numerous channels that pass between cell walls of adjacent plant cells, connect their cytoplasm, and enable materials to be transported from cell to cell, and thus throughout the plant ([link](#)).

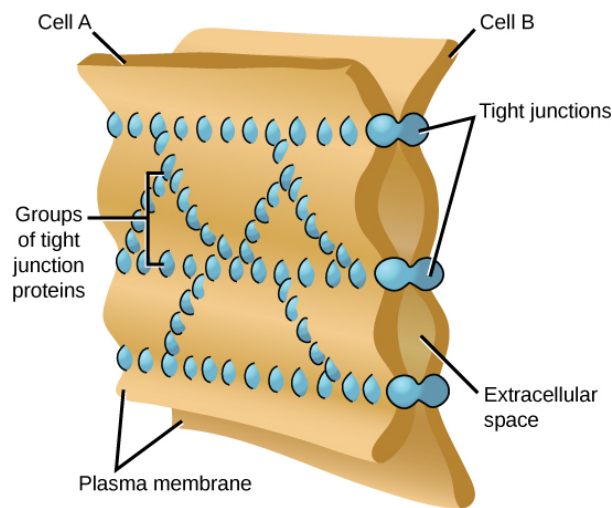


A plasmodesma is a channel between the cell walls of two adjacent plant cells.

Plasmodesmata allow materials to pass from the cytoplasm of one plant cell to the cytoplasm of an adjacent cell.

Tight Junctions

A **tight junction** is a watertight seal between two adjacent animal cells ([link](#)). The cells are held tightly against each other by proteins (predominantly two proteins called claudins and occludins).



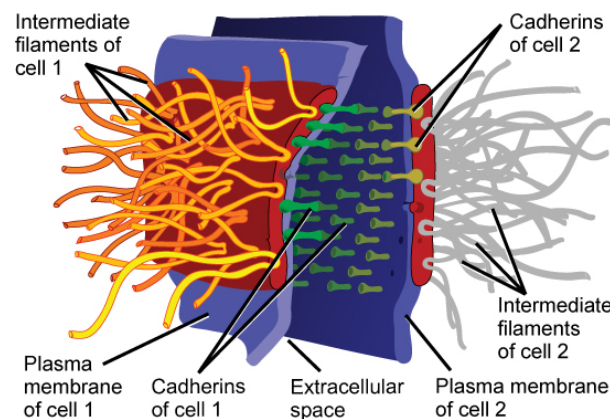
Tight junctions form watertight connections between adjacent animal cells. Proteins create tight junction adherence.

(credit: modification of work by Mariana Ruiz Villareal)

This tight adherence prevents materials from leaking between the cells; tight junctions are typically found in epithelial tissues that line internal organs and cavities, and comprise most of the skin. For example, the tight junctions of the epithelial cells lining your urinary bladder prevent urine from leaking out into the extracellular space.

Desmosomes

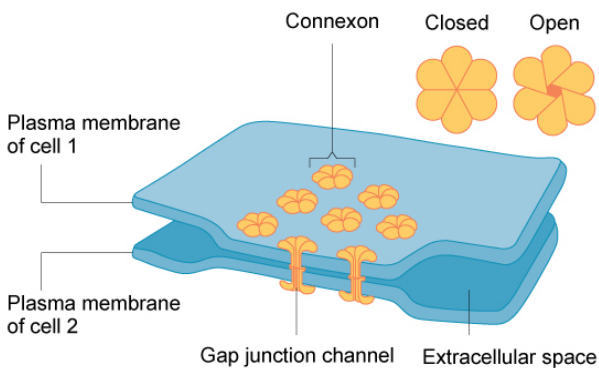
Also found only in animal cells are **desmosomes**, which act like spot welds between adjacent epithelial cells ([\[link\]](#)). Short proteins called cadherins in the plasma membrane connect to intermediate filaments to create desmosomes. The cadherins join two adjacent cells together and maintain the cells in a sheet-like formation in organs and tissues that stretch, like the skin, heart, and muscles.



A desmosome forms a very strong spot weld between cells. It is created by the linkage of cadherins and intermediate filaments. (credit: modification of work by Mariana Ruiz Villareal)

Gap Junctions

Gap junctions in animal cells are like plasmodesmata in plant cells in that they are channels between adjacent cells that allow for the transport of ions, nutrients, and other substances that enable cells to communicate ([\[link\]](#)). Structurally, however, gap junctions and plasmodesmata differ.



A gap junction is a protein-lined pore that allows water and small molecules to pass between adjacent animal cells. (credit: modification of work by Mariana Ruiz Villareal)

Gap junctions develop when a set of six proteins (called connexins) in the plasma membrane arrange themselves in an elongated donut-like configuration called a connexon. When the pores (“doughnut holes”) of connexons in adjacent animal cells align, a channel between the two cells forms. Gap junctions are particularly important in cardiac muscle: The electrical signal for the muscle to contract is passed efficiently through gap junctions, allowing the heart muscle cells to contract in tandem.

Note:

Link to Learning



To conduct a virtual microscopy lab and review the parts of a cell, work through the steps of this [interactive assignment](#).

Section Summary

Animal cells communicate via their extracellular matrices and are connected to each other via tight junctions, desmosomes, and gap junctions. Plant cells are connected and communicate with each other via plasmodesmata.

When protein receptors on the surface of the plasma membrane of an animal cell bind to a substance in the extracellular matrix, a chain of reactions begins that changes activities taking place within the cell. Plasmodesmata are channels between adjacent plant cells, while gap junctions are channels between adjacent animal cells. However, their structures are quite different. A tight junction is a watertight seal between two adjacent cells, while a desmosome acts like a spot weld.

Review Questions

Exercise:

Problem: Which of the following are found only in plant cells?

- a. gap junctions
- b. desmosomes
- c. plasmodesmata

d. tight junctions

Solution:

C

Exercise:

Problem:

The key components of desmosomes are cadherins and _____.

- a. actin
- b. microfilaments
- c. intermediate filaments
- d. microtubules

Solution:

C

Free Response

Exercise:

Problem:

How does the structure of a plasmodesma differ from that of a gap junction?

Solution:

They differ because plant cell walls are rigid. Plasmodesmata, which a plant cell needs for transportation and communication, are able to allow movement of really large molecules. Gap junctions are necessary in animal cells for transportation and communication.

Exercise:

Problem: Explain how the extracellular matrix functions.

Solution:

The extracellular matrix functions in support and attachment for animal tissues. It also functions in the healing and growth of the tissue.

Glossary

desmosome

linkages between adjacent epithelial cells that form when cadherins in the plasma membrane attach to intermediate filaments

extracellular matrix

material (primarily collagen, glycoproteins, and proteoglycans) secreted from animal cells that provides mechanical protection and anchoring for the cells in the tissue

gap junction

channel between two adjacent animal cells that allows ions, nutrients, and low molecular weight substances to pass between cells, enabling the cells to communicate

plasmodesma

(plural = plasmodesmata) channel that passes between the cell walls of adjacent plant cells, connects their cytoplasm, and allows materials to be transported from cell to cell

tight junction

firm seal between two adjacent animal cells created by protein adherence

Introduction

class="introduction"

Despite its
seeming
hustle and
bustle,
Grand
Central
Station
functions
with a high
level of
organization
: People and
objects
move from
one location
to another,
they cross or
are
contained
within
certain
boundaries,
and they
provide a
constant
flow as part
of larger
activity.
Analogously
, a plasma
membrane's
functions
involve
movement

within the
cell and
across
boundaries
in the
process of
intracellular
and
intercellular
activities.

(credit:
modification
of work by
Randy
Le'Moine)



The plasma membrane, which is also called the cell membrane, has many functions, but the most basic one is to define the borders of the cell and keep the cell functional. The plasma membrane is selectively permeable. This means that the membrane allows some materials to freely enter or leave the cell, while other materials cannot move freely, but require the use

of a specialized structure, and occasionally, even energy investment for crossing.

Components and Structure

By the end of this section, you will be able to:

- Understand the fluid mosaic model of cell membranes
- Describe the functions of phospholipids, proteins, and carbohydrates in membranes
- Discuss membrane fluidity

A cell's plasma membrane defines the cell, outlines its borders, and determines the nature of its interaction with its environment (see [\[link\]](#) for a summary). Cells exclude some substances, take in others, and excrete still others, all in controlled quantities. The plasma membrane must be very flexible to allow certain cells, such as red blood cells and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious functions of a plasma membrane. In addition, the surface of the plasma membrane carries markers that allow cells to recognize one another, which is vital for tissue and organ formation during early development, and which later plays a role in the “self” versus “non-self” distinction of the immune response.

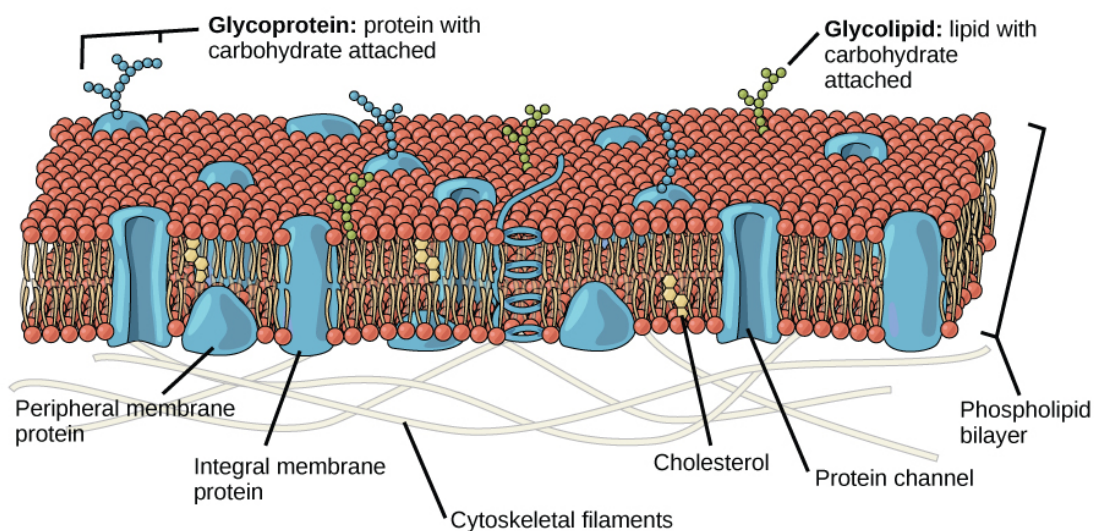
Among the most sophisticated functions of the plasma membrane is the ability to transmit signals by means of complex, integral proteins known as receptors. These proteins act both as receivers of extracellular inputs and as activators of intracellular processes. These membrane receptors provide extracellular attachment sites for effectors like hormones and growth factors, and they activate intracellular response cascades when their effectors are bound. Occasionally, receptors are hijacked by viruses (HIV, human immunodeficiency virus, is one example) that use them to gain entry into cells, and at times, the genes encoding receptors become mutated, causing the process of signal transduction to malfunction with disastrous consequences.

Fluid Mosaic Model

The existence of the plasma membrane was identified in the 1890s, and its chemical components were identified in 1915. The principal components identified at that time were lipids and proteins. The first widely accepted

model of the plasma membrane's structure was proposed in 1935 by Hugh Davson and James Danielli; it was based on the “railroad track” appearance of the plasma membrane in early electron micrographs. They theorized that the structure of the plasma membrane resembles a sandwich, with protein being analogous to the bread, and lipids being analogous to the filling. In the 1950s, advances in microscopy, notably transmission electron microscopy (TEM), allowed researchers to see that the core of the plasma membrane consisted of a double, rather than a single, layer. A new model that better explains both the microscopic observations and the function of that plasma membrane was proposed by S.J. Singer and Garth L. Nicolson in 1972.

The explanation proposed by Singer and Nicolson is called the **fluid mosaic model**. The model has evolved somewhat over time, but it still best accounts for the structure and functions of the plasma membrane as we now understand them. The fluid mosaic model describes the structure of the plasma membrane as a mosaic of components—including phospholipids, cholesterol, proteins, and carbohydrates—that gives the membrane a fluid character. Plasma membranes range from 5 to 10 nm in thickness. For comparison, human red blood cells, visible via light microscopy, are approximately 8 μm wide, or approximately 1,000 times wider than a plasma membrane. The membrane does look a bit like a sandwich ([\[link\]](#)).



The fluid mosaic model of the plasma membrane describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the outward-facing surface of the membrane.

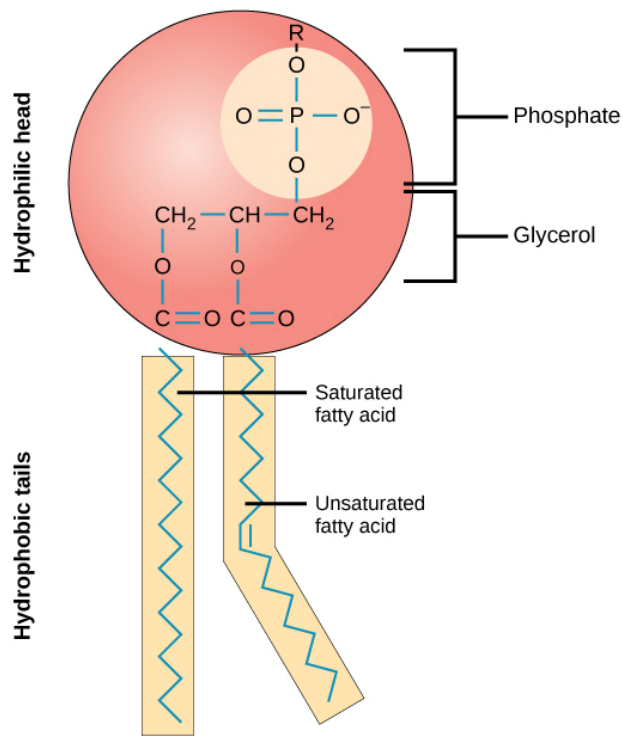
The principal components of a plasma membrane are lipids (phospholipids and cholesterol), proteins, and carbohydrates attached to some of the lipids and some of the proteins. A phospholipid is a molecule consisting of glycerol, two fatty acids, and a phosphate-linked head group. Cholesterol, another lipid composed of four fused carbon rings, is found alongside the phospholipids in the core of the membrane. The proportions of proteins, lipids, and carbohydrates in the plasma membrane vary with cell type, but for a typical human cell, protein accounts for about 50 percent of the composition by mass, lipids (of all types) account for about 40 percent of the composition by mass, with the remaining 10 percent of the composition by mass being carbohydrates. However, the concentration of proteins and lipids varies with different cell membranes. For example, myelin, an outgrowth of the membrane of specialized cells that insulates the axons of the peripheral nerves, contains only 18 percent protein and 76 percent lipid. The mitochondrial inner membrane contains 76 percent protein and only 24 percent lipid. The plasma membrane of human red blood cells is 30 percent lipid. Carbohydrates are present only on the exterior surface of the plasma membrane and are attached to proteins, forming **glycoproteins**, or attached to lipids, forming **glycolipids**.

Phospholipids

The main fabric of the membrane is composed of amphiphilic, phospholipid molecules. The **hydrophilic** or “water-loving” areas of these molecules (which look like a collection of balls in an artist’s rendition of the model) ([\[link\]](#)) are in contact with the aqueous fluid both inside and outside the cell. **Hydrophobic**, or water-hating molecules, tend to be non-polar. They interact with other non-polar molecules in chemical reactions, but generally

do not interact with polar molecules. When placed in water, hydrophobic molecules tend to form a ball or cluster. The hydrophilic regions of the phospholipids tend to form hydrogen bonds with water and other polar molecules on both the exterior and interior of the cell. Thus, the membrane surfaces that face the interior and exterior of the cell are hydrophilic. In contrast, the interior of the cell membrane is hydrophobic and will not interact with water. Therefore, phospholipids form an excellent two-layer cell membrane that separates fluid within the cell from the fluid outside of the cell.

A phospholipid molecule ([\[link\]](#)) consists of a three-carbon glycerol backbone with two fatty acid molecules attached to carbons 1 and 2, and a phosphate-containing group attached to the third carbon. This arrangement gives the overall molecule an area described as its head (the phosphate-containing group), which has a polar character or negative charge, and an area called the tail (the fatty acids), which has no charge. The head can form hydrogen bonds, but the tail cannot. A molecule with this arrangement of a positively or negatively charged area and an uncharged, or non-polar, area is referred to as **amphiphilic** or “dual-loving.”

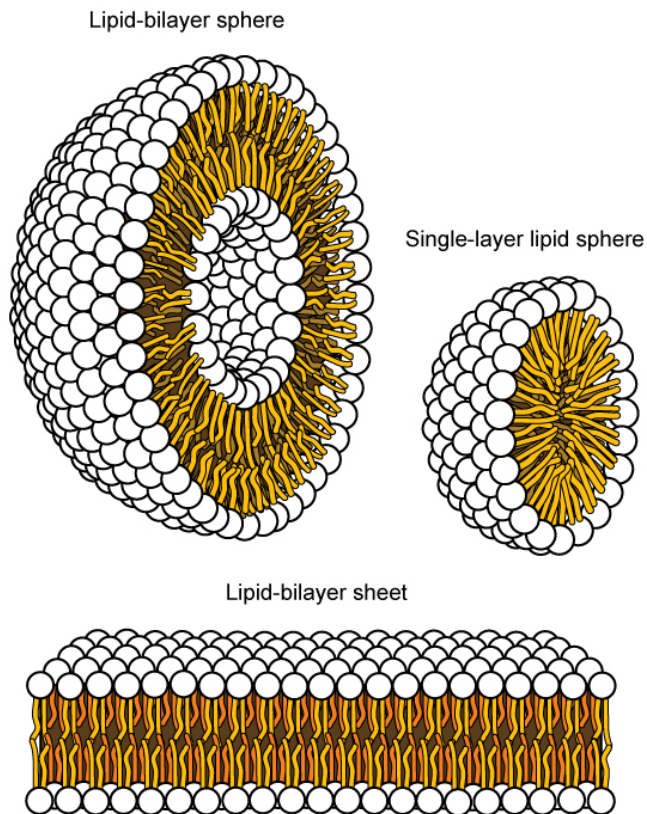


This phospholipid molecule is composed of a hydrophilic head and two hydrophobic tails. The hydrophilic head group consists of a phosphate-containing group attached to a glycerol molecule.

The hydrophobic tails, each containing either a saturated or an unsaturated fatty acid, are long hydrocarbon chains.

This characteristic is vital to the structure of a plasma membrane because, in water, phospholipids tend to become arranged with their hydrophobic tails facing each other and their hydrophilic heads facing out. In this way, they form a lipid bilayer—a barrier composed of a double layer of phospholipids that separates the water and other materials on one side of the barrier from the water and other materials on the other side. In fact, phospholipids heated in an aqueous solution tend to spontaneously form

small spheres or droplets (called micelles or liposomes), with their hydrophilic heads forming the exterior and their hydrophobic tails on the inside ([link](#)).

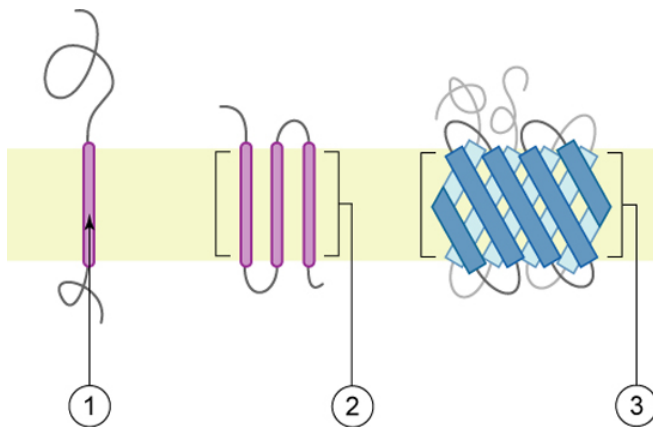


In an aqueous solution, phospholipids tend to arrange themselves with their polar heads facing outward and their hydrophobic tails facing inward.
(credit: modification of work by Mariana Ruiz Villareal)

Proteins

Proteins make up the second major component of plasma membranes.

Integral proteins (some specialized types are called integrins) are, as their name suggests, integrated completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the hydrophobic region of the phospholipid bilayer ([\[link\]](#)). Single-pass integral membrane proteins usually have a hydrophobic transmembrane segment that consists of 20–25 amino acids. Some span only part of the membrane—associating with a single layer—while others stretch from one side of the membrane to the other, and are exposed on either side. Some complex proteins are composed of up to 12 segments of a single protein, which are extensively folded and embedded in the membrane ([\[link\]](#)). This type of protein has a hydrophilic region or regions, and one or several mildly hydrophobic regions. This arrangement of regions of the protein tends to orient the protein alongside the phospholipids, with the hydrophobic region of the protein adjacent to the tails of the phospholipids and the hydrophilic region or regions of the protein protruding from the membrane and in contact with the cytosol or extracellular fluid.



Integral membranes proteins may have one or more alpha-helices that span the membrane (examples 1 and 2), or they may have beta-sheets that span the membrane (example 3). (credit: “Foobar”/Wikimedia Commons)

Peripheral proteins are found on the exterior and interior surfaces of membranes, attached either to integral proteins or to phospholipids. Peripheral proteins, along with integral proteins, may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell's recognition sites. These are sometimes referred to as “cell-specific” proteins. The body recognizes its own proteins and attacks foreign proteins associated with invasive pathogens.

Carbohydrates

Carbohydrates are the third major component of plasma membranes. They are always found on the exterior surface of cells and are bound either to proteins (forming glycoproteins) or to lipids (forming glycolipids) ([\[link\]](#)). These carbohydrate chains may consist of 2–60 monosaccharide units and can be either straight or branched. Along with peripheral proteins, carbohydrates form specialized sites on the cell surface that allow cells to recognize each other. These sites have unique patterns that allow the cell to be recognized, much the way that the facial features unique to each person allow him or her to be recognized. This recognition function is very important to cells, as it allows the immune system to differentiate between body cells (called “self”) and foreign cells or tissues (called “non-self”). Similar types of glycoproteins and glycolipids are found on the surfaces of viruses and may change frequently, preventing immune cells from recognizing and attacking them.

These carbohydrates on the exterior surface of the cell—the carbohydrate components of both glycoproteins and glycolipids—are collectively referred to as the glycocalyx (meaning “sugar coating”). The glycocalyx is highly hydrophilic and attracts large amounts of water to the surface of the cell. This aids in the interaction of the cell with its watery environment and in the cell's ability to obtain substances dissolved in the water. As discussed above, the glycocalyx is also important for cell identification, self/non-self

determination, and embryonic development, and is used in cell-cell attachments to form tissues.

Note:

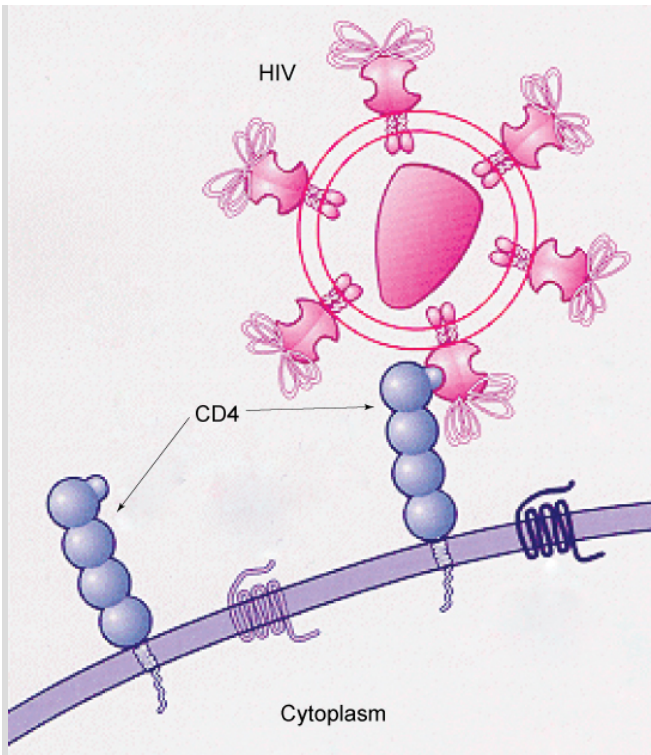
Evolution Connection

How Viruses Infect Specific Organs

Glycoprotein and glycolipid patterns on the surfaces of cells give many viruses an opportunity for infection. HIV and hepatitis viruses infect only specific organs or cells in the human body. HIV is able to penetrate the plasma membranes of a subtype of lymphocytes called T-helper cells, as well as some monocytes and central nervous system cells. The hepatitis virus attacks liver cells.

These viruses are able to invade these cells, because the cells have binding sites on their surfaces that are specific to and compatible with certain viruses ([link](#)). Other recognition sites on the virus's surface interact with the human immune system, prompting the body to produce antibodies.

Antibodies are made in response to the antigens or proteins associated with invasive pathogens, or in response to foreign cells, such as might occur with an organ transplant. These same sites serve as places for antibodies to attach and either destroy or inhibit the activity of the virus. Unfortunately, these recognition sites on HIV change at a rapid rate because of mutations, making the production of an effective vaccine against the virus very difficult, as the virus evolves and adapts. A person infected with HIV will quickly develop different populations, or variants, of the virus that are distinguished by differences in these recognition sites. This rapid change of surface markers decreases the effectiveness of the person's immune system in attacking the virus, because the antibodies will not recognize the new variations of the surface patterns. In the case of HIV, the problem is compounded by the fact that the virus specifically infects and destroys cells involved in the immune response, further incapacitating the host.



HIV binds to the CD4 receptor, a glycoprotein on the surfaces of T cells. (credit: modification of work by NIH, NIAID)

Membrane Fluidity

The mosaic characteristic of the membrane, described in the fluid mosaic model, helps to illustrate its nature. The integral proteins and lipids exist in the membrane as separate but loosely attached molecules. These resemble the separate, multicolored tiles of a mosaic picture, and they float, moving somewhat with respect to one another. The membrane is not like a balloon, however, that can expand and contract; rather, it is fairly rigid and can burst if penetrated or if a cell takes in too much water. However, because of its mosaic nature, a very fine needle can easily penetrate a plasma membrane

without causing it to burst, and the membrane will flow and self-seal when the needle is extracted.

The mosaic characteristics of the membrane explain some but not all of its fluidity. There are two other factors that help maintain this fluid characteristic. One factor is the nature of the phospholipids themselves. In their saturated form, the fatty acids in phospholipid tails are saturated with bound hydrogen atoms. There are no double bonds between adjacent carbon atoms. This results in tails that are relatively straight. In contrast, unsaturated fatty acids do not contain a maximal number of hydrogen atoms, but they do contain some double bonds between adjacent carbon atoms; a double bond results in a bend in the string of carbons of approximately 30 degrees ([\[link\]](#)).

Thus, if saturated fatty acids, with their straight tails, are compressed by decreasing temperatures, they press in on each other, making a dense and fairly rigid membrane. If unsaturated fatty acids are compressed, the “kinks” in their tails elbow adjacent phospholipid molecules away, maintaining some space between the phospholipid molecules. This “elbow room” helps to maintain fluidity in the membrane at temperatures at which membranes with saturated fatty acid tails in their phospholipids would “freeze” or solidify. The relative fluidity of the membrane is particularly important in a cold environment. A cold environment tends to compress membranes composed largely of saturated fatty acids, making them less fluid and more susceptible to rupturing. Many organisms (fish are one example) are capable of adapting to cold environments by changing the proportion of unsaturated fatty acids in their membranes in response to the lowering of the temperature.

Note:

Link to Learning



Visit this [site](#) to see animations of the fluidity and mosaic quality of membranes.

Animals have an additional membrane constituent that assists in maintaining fluidity. Cholesterol, which lies alongside the phospholipids in the membrane, tends to dampen the effects of temperature on the membrane. Thus, this lipid functions as a buffer, preventing lower temperatures from inhibiting fluidity and preventing increased temperatures from increasing fluidity too much. Thus, cholesterol extends, in both directions, the range of temperature in which the membrane is appropriately fluid and consequently functional. Cholesterol also serves other functions, such as organizing clusters of transmembrane proteins into lipid rafts.

The Components and Functions of the Plasma Membrane	
Component	Location
Phospholipid	Main fabric of the membrane
Cholesterol	Attached between phospholipids and between the two phospholipid layers

The Components and Functions of the Plasma Membrane	
Component	Location
Integral proteins (for example, integrins)	Embedded within the phospholipid layer(s). May or may not penetrate through both layers
Peripheral proteins	On the inner or outer surface of the phospholipid bilayer; not embedded within the phospholipids
Carbohydrates (components of glycoproteins and glycolipids)	Generally attached to proteins on the outside membrane layer

Note:

Career Connection

Immunologist

The variations in peripheral proteins and carbohydrates that affect a cell's recognition sites are of prime interest in immunology. These changes are taken into consideration in vaccine development. Many infectious diseases, such as smallpox, polio, diphtheria, and tetanus, were conquered by the use of vaccines.

Immunologists are the physicians and scientists who research and develop vaccines, as well as treat and study allergies or other immune problems. Some immunologists study and treat autoimmune problems (diseases in which a person's immune system attacks his or her own cells or tissues, such as lupus) and immunodeficiencies, whether acquired (such as acquired immunodeficiency syndrome, or AIDS) or hereditary (such as severe combined immunodeficiency, or SCID). Immunologists are called in to help treat organ transplantation patients, who must have their immune systems suppressed so that their bodies will not reject a transplanted organ.

Some immunologists work to understand natural immunity and the effects of a person's environment on it. Others work on questions about how the immune system affects diseases such as cancer. In the past, the importance of having a healthy immune system in preventing cancer was not at all understood.

To work as an immunologist, a PhD or MD is required. In addition, immunologists undertake at least 2–3 years of training in an accredited program and must pass an examination given by the American Board of Allergy and Immunology. Immunologists must possess knowledge of the functions of the human body as they relate to issues beyond immunization, and knowledge of pharmacology and medical technology, such as medications, therapies, test materials, and surgical procedures.

Section Summary

The modern understanding of the plasma membrane is referred to as the fluid mosaic model. The plasma membrane is composed of a bilayer of phospholipids, with their hydrophobic, fatty acid tails in contact with each other. The landscape of the membrane is studded with proteins, some of which span the membrane. Some of these proteins serve to transport materials into or out of the cell. Carbohydrates are attached to some of the proteins and lipids on the outward-facing surface of the membrane, forming complexes that function to identify the cell to other cells. The fluid nature of the membrane is due to temperature, the configuration of the fatty acid tails (some kinked by double bonds), the presence of cholesterol embedded in the membrane, and the mosaic nature of the proteins and protein-carbohydrate combinations, which are not firmly fixed in place. Plasma membranes enclose and define the borders of cells, but rather than being a static bag, they are dynamic and constantly in flux.

Review Questions

Exercise:

Problem:

Which plasma membrane component can be either found on its surface or embedded in the membrane structure?

- a. protein
- b. cholesterol
- c. carbohydrate
- d. phospholipid

Solution:

A

Exercise:**Problem:**

Which characteristic of a phospholipid contributes to the fluidity of the membrane?

- a. its head
- b. cholesterol
- c. a saturated fatty acid tail
- d. double bonds in the fatty acid tail

Solution:

D

Exercise:**Problem:**

What is the primary function of carbohydrates attached to the exterior of cell membranes?

- a. identification of the cell

- b. flexibility of the membrane
- c. strengthening the membrane
- d. channels through membrane

Solution:

A

Free Response

Exercise:

Problem:

Why is it advantageous for the cell membrane to be fluid in nature?

Solution:

The fluid characteristic of the cell membrane allows greater flexibility to the cell than it would if the membrane were rigid. It also allows the motion of membrane components, required for some types of membrane transport.

Exercise:

Problem:

Why do phospholipids tend to spontaneously orient themselves into something resembling a membrane?

Solution:

The hydrophobic, nonpolar regions must align with each other in order for the structure to have minimal potential energy and, consequently, higher stability. The fatty acid tails of the phospholipids cannot mix with water, but the phosphate “head” of the molecule can. Thus, the head orients to water, and the tail to other lipids.

Glossary

amphiphilic

molecule possessing a polar or charged area and a nonpolar or uncharged area capable of interacting with both hydrophilic and hydrophobic environments

fluid mosaic model

describes the structure of the plasma membrane as a mosaic of components including phospholipids, cholesterol, proteins, glycoproteins, and glycolipids (sugar chains attached to proteins or lipids, respectively), resulting in a fluid character (fluidity)

glycolipid

combination of carbohydrates and lipids

glycoprotein

combination of carbohydrates and proteins

hydrophilic

molecule with the ability to bond with water; “water-loving”

hydrophobic

molecule that does not have the ability to bond with water; “water-hating”

integral protein

protein integrated into the membrane structure that interacts extensively with the hydrocarbon chains of membrane lipids and often spans the membrane; these proteins can be removed only by the disruption of the membrane by detergents

peripheral protein

protein found at the surface of a plasma membrane either on its exterior or interior side; these proteins can be removed (washed off of the membrane) by a high-salt wash

Passive Transport

By the end of this section, you will be able to:

- Explain why and how passive transport occurs
- Understand the processes of osmosis and diffusion
- Define tonicity and describe its relevance to passive transport

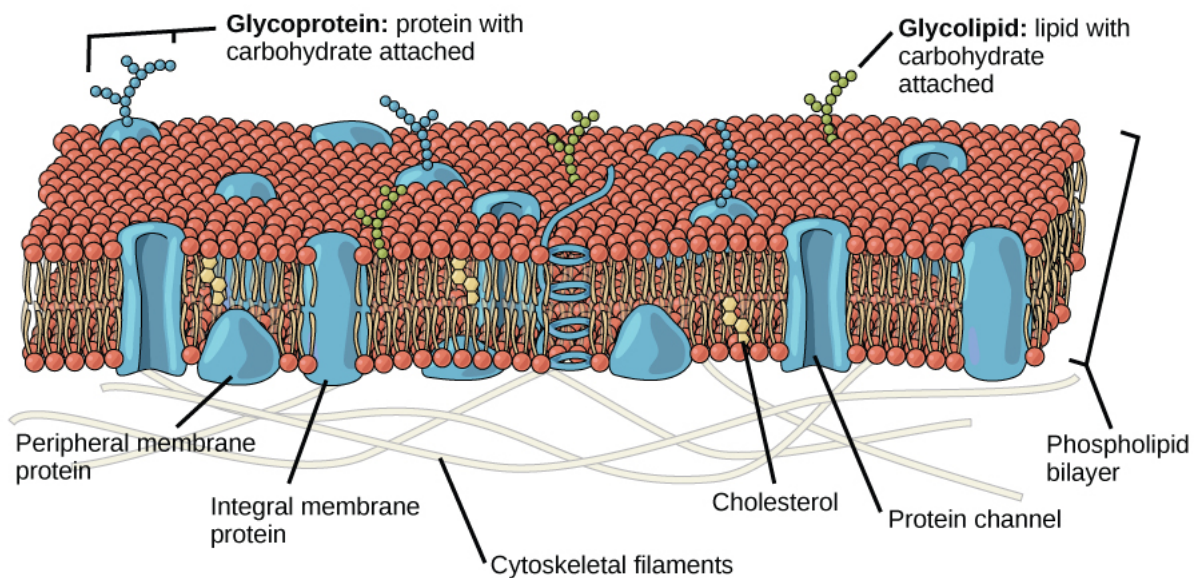
Plasma membranes must allow certain substances to enter and leave a cell, and prevent some harmful materials from entering and some essential materials from leaving. In other words, plasma membranes are **selectively permeable**—they allow some substances to pass through, but not others. If they were to lose this selectivity, the cell would no longer be able to sustain itself, and it would be destroyed. Some cells require larger amounts of specific substances than do other cells; they must have a way of obtaining these materials from extracellular fluids. This may happen passively, as certain materials move back and forth, or the cell may have special mechanisms that facilitate transport. Some materials are so important to a cell that it spends some of its energy, hydrolyzing adenosine triphosphate (ATP), to obtain these materials. Red blood cells use some of their energy doing just that. All cells spend the majority of their energy to maintain an imbalance of sodium and potassium ions between the interior and exterior of the cell.

The most direct forms of membrane transport are passive. **Passive transport** is a naturally occurring phenomenon and does not require the cell to exert any of its energy to accomplish the movement. In passive transport, substances move from an area of higher concentration to an area of lower concentration. A physical space in which there is a range of concentrations of a single substance is said to have a **concentration gradient**.

Selective Permeability

Plasma membranes are asymmetric: the interior of the membrane is not identical to the exterior of the membrane. In fact, there is a considerable difference between the array of phospholipids and proteins between the two leaflets that form a membrane. On the interior of the membrane, some proteins serve to anchor the membrane to fibers of the cytoskeleton. There

are peripheral proteins on the exterior of the membrane that bind elements of the extracellular matrix. Carbohydrates, attached to lipids or proteins, are also found on the exterior surface of the plasma membrane. These carbohydrate complexes help the cell bind substances that the cell needs in the extracellular fluid. This adds considerably to the selective nature of plasma membranes ([\[link\]](#)).



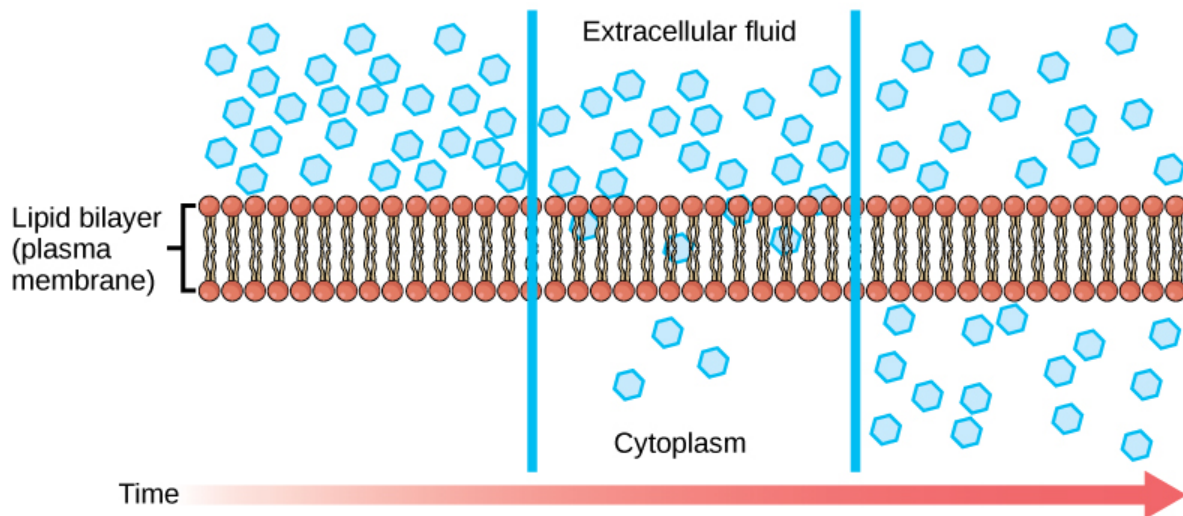
The exterior surface of the plasma membrane is not identical to the interior surface of the same membrane.

Recall that plasma membranes are amphiphilic: They have hydrophilic and hydrophobic regions. This characteristic helps the movement of some materials through the membrane and hinders the movement of others. Lipid-soluble material with a low molecular weight can easily slip through the hydrophobic lipid core of the membrane. Substances such as the fat-soluble vitamins A, D, E, and K readily pass through the plasma membranes in the digestive tract and other tissues. Fat-soluble drugs and hormones also gain easy entry into cells and are readily transported into the body's tissues and organs. Molecules of oxygen and carbon dioxide have no charge and so pass through membranes by simple diffusion.

Polar substances present problems for the membrane. While some polar molecules connect easily with the outside of a cell, they cannot readily pass through the lipid core of the plasma membrane. Additionally, while small ions could easily slip through the spaces in the mosaic of the membrane, their charge prevents them from doing so. Ions such as sodium, potassium, calcium, and chloride must have special means of penetrating plasma membranes. Simple sugars and amino acids also need help with transport across plasma membranes, achieved by various transmembrane proteins (channels).

Diffusion

Diffusion is a passive process of transport. A single substance tends to move from an area of high concentration to an area of low concentration until the concentration is equal across a space. You are familiar with diffusion of substances through the air. For example, think about someone opening a bottle of ammonia in a room filled with people. The ammonia gas is at its highest concentration in the bottle; its lowest concentration is at the edges of the room. The ammonia vapor will diffuse, or spread away, from the bottle, and gradually, more and more people will smell the ammonia as it spreads. Materials move within the cell's cytosol by diffusion, and certain materials move through the plasma membrane by diffusion ([\[link\]](#)). Diffusion expends no energy. On the contrary, concentration gradients are a form of potential energy, dissipated as the gradient is eliminated.



Diffusion through a permeable membrane moves a substance from an area of high concentration (extracellular fluid, in this case) down its concentration gradient (into the cytoplasm). (credit: modification of work by Mariana Ruiz Villareal)

Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of the concentration gradients of other materials. In addition, each substance will diffuse according to that gradient. Within a system, there will be different rates of diffusion of the different substances in the medium.

Factors That Affect Diffusion

Molecules move constantly in a random manner, at a rate that depends on their mass, their environment, and the amount of thermal energy they possess, which in turn is a function of temperature. This movement accounts for the diffusion of molecules through whatever medium in which they are localized. A substance will tend to move into any space available to it until it is evenly distributed throughout it. After a substance has diffused completely through a space, removing its concentration gradient, molecules will still move around in the space, but there will be no *net*

movement of the number of molecules from one area to another. This lack of a concentration gradient in which there is no net movement of a substance is known as dynamic equilibrium. While diffusion will go forward in the presence of a concentration gradient of a substance, several factors affect the rate of diffusion.

- **Extent of the concentration gradient:** The greater the difference in concentration, the more rapid the diffusion. The closer the distribution of the material gets to equilibrium, the slower the rate of diffusion becomes.
- **Mass of the molecules diffusing:** Heavier molecules move more slowly; therefore, they diffuse more slowly. The reverse is true for lighter molecules.
- **Temperature:** Higher temperatures increase the energy and therefore the movement of the molecules, increasing the rate of diffusion. Lower temperatures decrease the energy of the molecules, thus decreasing the rate of diffusion.
- **Solvent density:** As the density of a solvent increases, the rate of diffusion decreases. The molecules slow down because they have a more difficult time getting through the denser medium. If the medium is less dense, diffusion increases. Because cells primarily use diffusion to move materials within the cytoplasm, any increase in the cytoplasm's density will inhibit the movement of the materials. An example of this is a person experiencing dehydration. As the body's cells lose water, the rate of diffusion decreases in the cytoplasm, and the cells' functions deteriorate. Neurons tend to be very sensitive to this effect. Dehydration frequently leads to unconsciousness and possibly coma because of the decrease in diffusion rate within the cells.
- **Solubility:** As discussed earlier, nonpolar or lipid-soluble materials pass through plasma membranes more easily than polar materials, allowing a faster rate of diffusion.
- **Surface area and thickness of the plasma membrane:** Increased surface area increases the rate of diffusion, whereas a thicker membrane reduces it.
- **Distance travelled:** The greater the distance that a substance must travel, the slower the rate of diffusion. This places an upper limitation

on cell size. A large, spherical cell will die because nutrients or waste cannot reach or leave the center of the cell, respectively. Therefore, cells must either be small in size, as in the case of many prokaryotes, or be flattened, as with many single-celled eukaryotes.

A variation of diffusion is the process of filtration. In filtration, material moves according to its concentration gradient through a membrane; sometimes the rate of diffusion is enhanced by pressure, causing the substances to filter more rapidly. This occurs in the kidney, where blood pressure forces large amounts of water and accompanying dissolved substances, or **solutes**, out of the blood and into the renal tubules. The rate of diffusion in this instance is almost totally dependent on pressure. One of the effects of high blood pressure is the appearance of protein in the urine, which is “squeezed through” by the abnormally high pressure.

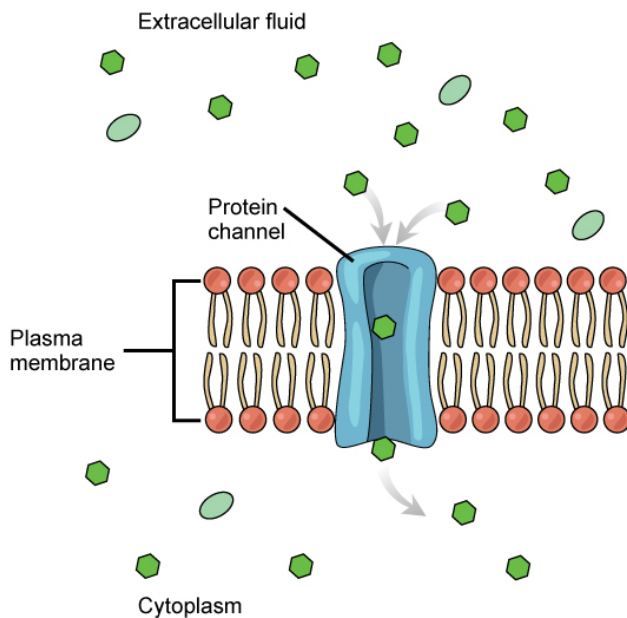
Facilitated transport

In **facilitated transport**, also called facilitated diffusion, materials diffuse across the plasma membrane with the help of membrane proteins. A concentration gradient exists that would allow these materials to diffuse into the cell without expending cellular energy. However, these materials are ions or polar molecules that are repelled by the hydrophobic parts of the cell membrane. Facilitated transport proteins shield these materials from the repulsive force of the membrane, allowing them to diffuse into the cell.

The material being transported is first attached to protein or glycoprotein receptors on the exterior surface of the plasma membrane. This allows the material that is needed by the cell to be removed from the extracellular fluid. The substances are then passed to specific integral proteins that facilitate their passage. Some of these integral proteins are collections of beta pleated sheets that form a pore or channel through the phospholipid bilayer. Others are carrier proteins which bind with the substance and aid its diffusion through the membrane.

Channels

The integral proteins involved in facilitated transport are collectively referred to as **transport proteins**, and they function as either channels for the material or carriers. In both cases, they are transmembrane proteins. Channels are specific for the substance that is being transported. **Channel proteins** have hydrophilic domains exposed to the intracellular and extracellular fluids; they additionally have a hydrophilic channel through their core that provides a hydrated opening through the membrane layers ([link](#)). Passage through the channel allows polar compounds to avoid the nonpolar central layer of the plasma membrane that would otherwise slow or prevent their entry into the cell. **Aquaporins** are channel proteins that allow water to pass through the membrane at a very high rate.

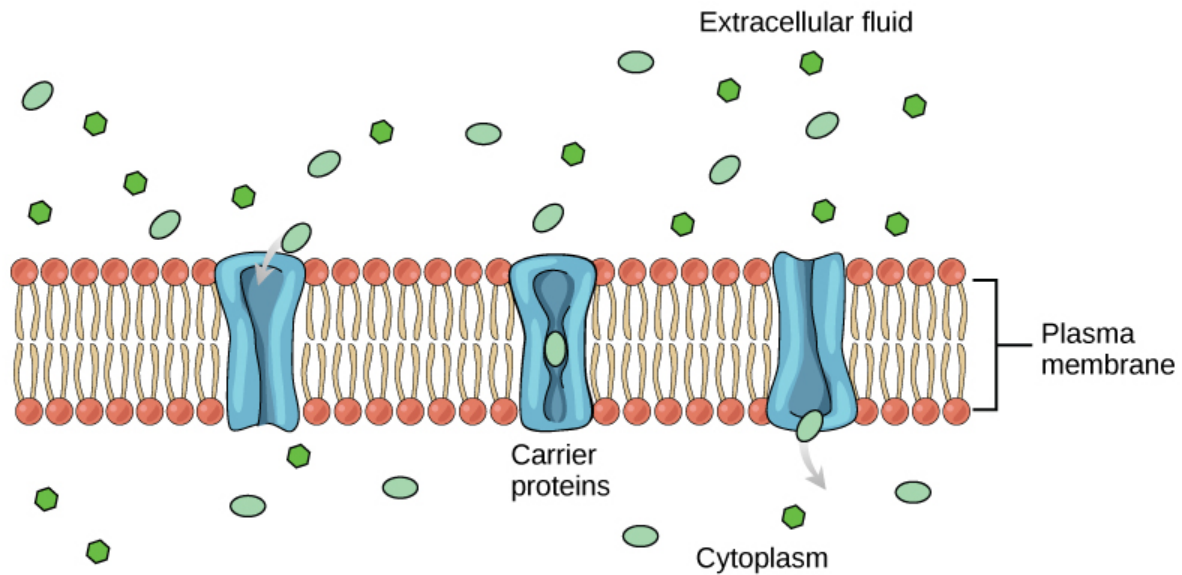


Facilitated transport moves substances down their concentration gradients. They may cross the plasma membrane with the aid of channel proteins. (credit: modification of work by Mariana Ruiz Villareal)

Channel proteins are either open at all times or they are “gated,” which controls the opening of the channel. The attachment of a particular ion to the channel protein may control the opening, or other mechanisms or substances may be involved. In some tissues, sodium and chloride ions pass freely through open channels, whereas in other tissues a gate must be opened to allow passage. An example of this occurs in the kidney, where both forms of channels are found in different parts of the renal tubules. Cells involved in the transmission of electrical impulses, such as nerve and muscle cells, have gated channels for sodium, potassium, and calcium in their membranes. Opening and closing of these channels changes the relative concentrations on opposing sides of the membrane of these ions, resulting in the facilitation of electrical transmission along membranes (in the case of nerve cells) or in muscle contraction (in the case of muscle cells).

Carrier Proteins

Another type of protein embedded in the plasma membrane is a **carrier protein**. This aptly named protein binds a substance and, in doing so, triggers a change of its own shape, moving the bound molecule from the outside of the cell to its interior ([link](#)); depending on the gradient, the material may move in the opposite direction. Carrier proteins are typically specific for a single substance. This selectivity adds to the overall selectivity of the plasma membrane. The exact mechanism for the change of shape is poorly understood. Proteins can change shape when their hydrogen bonds are affected, but this may not fully explain this mechanism. Each carrier protein is specific to one substance, and there are a finite number of these proteins in any membrane. This can cause problems in transporting enough of the material for the cell to function properly. When all of the proteins are bound to their ligands, they are saturated and the rate of transport is at its maximum. Increasing the concentration gradient at this point will not result in an increased rate of transport.



Some substances are able to move down their concentration gradient across the plasma membrane with the aid of carrier proteins. Carrier proteins change shape as they move molecules across the membrane.
(credit: modification of work by Mariana Ruiz Villareal)

An example of this process occurs in the kidney. Glucose, water, salts, ions, and amino acids needed by the body are filtered in one part of the kidney. This filtrate, which includes glucose, is then reabsorbed in another part of the kidney. Because there are only a finite number of carrier proteins for glucose, if more glucose is present than the proteins can handle, the excess is not transported and it is excreted from the body in the urine. In a diabetic individual, this is described as “spilling glucose into the urine.” A different group of carrier proteins called glucose transport proteins, or GLUTs, are involved in transporting glucose and other hexose sugars through plasma membranes within the body.

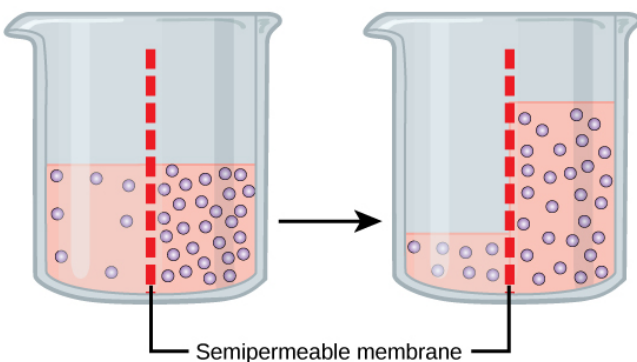
Channel and carrier proteins transport material at different rates. Channel proteins transport much more quickly than do carrier proteins. Channel proteins facilitate diffusion at a rate of tens of millions of molecules per second, whereas carrier proteins work at a rate of a thousand to a million molecules per second.

Osmosis

Osmosis is the movement of water through a semipermeable membrane according to the concentration gradient of water across the membrane, which is inversely proportional to the concentration of solutes. While diffusion transports material across membranes and within cells, osmosis transports *only water* across a membrane and the membrane limits the diffusion of solutes in the water. Not surprisingly, the aquaporins that facilitate water movement play a large role in osmosis, most prominently in red blood cells and the membranes of kidney tubules.

Mechanism

Osmosis is a special case of diffusion. Water, like other substances, moves from an area of high concentration to one of low concentration. An obvious question is what makes water move at all? Imagine a beaker with a semipermeable membrane separating the two sides or halves ([\[link\]](#)). On both sides of the membrane the water level is the same, but there are different concentrations of a dissolved substance, or **solute**, that cannot cross the membrane (otherwise the concentrations on each side would be balanced by the solute crossing the membrane). If the volume of the solution on both sides of the membrane is the same, but the concentrations of solute are different, then there are different amounts of water, the solvent, on either side of the membrane.



In osmosis, water always moves from an area of higher water concentration to one of lower concentration. In the diagram shown, the solute cannot pass through the selectively permeable membrane, but the water can.

To illustrate this, imagine two full glasses of water. One has a single teaspoon of sugar in it, whereas the second one contains one-quarter cup of sugar. If the total volume of the solutions in both cups is the same, which cup contains more water? Because the large amount of sugar in the second cup takes up much more space than the teaspoon of sugar in the first cup, the first cup has more water in it.

Returning to the beaker example, recall that it has a mixture of solutes on either side of the membrane. A principle of diffusion is that the molecules move around and will spread evenly throughout the medium if they can. However, only the material capable of getting through the membrane will diffuse through it. In this example, the solute cannot diffuse through the membrane, but the water can. Water has a concentration gradient in this system. Thus, water will diffuse down its concentration gradient, crossing the membrane to the side where it is less concentrated. This diffusion of water through the membrane—osmosis—will continue until the concentration gradient of water goes to zero or until the hydrostatic pressure of the water balances the osmotic pressure. Osmosis proceeds constantly in living systems.

Tonicity

Tonicity describes how an extracellular solution can change the volume of a cell by affecting osmosis. A solution's tonicity often directly correlates with the osmolarity of the solution. **Osmolarity** describes the total solute concentration of the solution. A solution with low osmolarity has a greater number of water molecules relative to the number of solute particles; a

solution with high osmolarity has fewer water molecules with respect to solute particles. In a situation in which solutions of two different osmolarities are separated by a membrane permeable to water, though not to the solute, water will move from the side of the membrane with lower osmolarity (and more water) to the side with higher osmolarity (and less water). This effect makes sense if you remember that the solute cannot move across the membrane, and thus the only component in the system that can move—the water—moves along its own concentration gradient. An important distinction that concerns living systems is that osmolarity measures the number of particles (which may be molecules) in a solution. Therefore, a solution that is cloudy with cells may have a lower osmolarity than a solution that is clear, if the second solution contains more dissolved molecules than there are cells.

Hypotonic Solutions

Three terms—hypotonic, isotonic, and hypertonic—are used to relate the osmolarity of a cell to the osmolarity of the extracellular fluid that contains the cells. In a **hypotonic** situation, the extracellular fluid has lower osmolarity than the fluid inside the cell, and water enters the cell. (In living systems, the point of reference is always the cytoplasm, so the prefix *hypo-* means that the extracellular fluid has a lower concentration of solutes, or a lower osmolarity, than the cell cytoplasm.) It also means that the extracellular fluid has a higher concentration of water in the solution than does the cell. In this situation, water will follow its concentration gradient and enter the cell.

Hypertonic Solutions

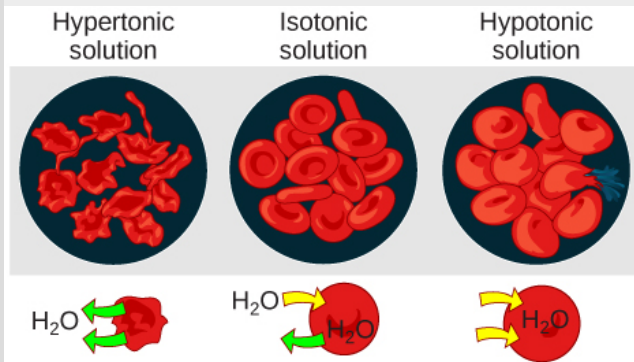
As for a **hypertonic** solution, the prefix *hyper-* refers to the extracellular fluid having a higher osmolarity than the cell's cytoplasm; therefore, the fluid contains less water than the cell does. Because the cell has a relatively higher concentration of water, water will leave the cell.

Isotonic Solutions

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. If the osmolarity of the cell matches that of the extracellular fluid, there will be no net movement of water into or out of the cell, although water will still move in and out. Blood cells and plant cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances ([link](#)).

Note:

Art Connection



Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions. (credit: Mariana Ruiz Villareal)

A doctor injects a patient with what the doctor thinks is an isotonic saline solution. The patient dies, and an autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

Note:

Link to Learning



For a video illustrating the process of diffusion in solutions, visit this [site](#).

Tonicity in Living Systems

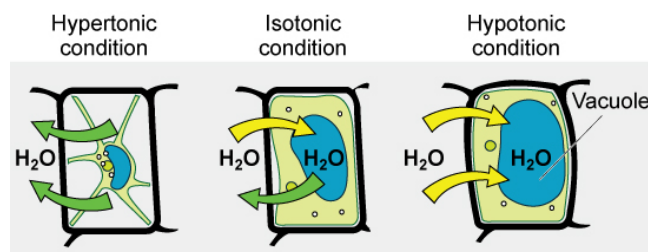
In a hypotonic environment, water enters a cell, and the cell swells. In an isotonic condition, the relative concentrations of solute and solvent are equal on both sides of the membrane. There is no net water movement; therefore, there is no change in the size of the cell. In a hypertonic solution, water leaves a cell and the cell shrinks. If either the hypo- or hyper-condition goes to excess, the cell's functions become compromised, and the cell may be destroyed.

A red blood cell will burst, or lyse, when it swells beyond the plasma membrane's capability to expand. Remember, the membrane resembles a mosaic, with discrete spaces between the molecules composing it. If the cell swells, and the spaces between the lipids and proteins become too large, the cell will break apart.

In contrast, when excessive amounts of water leave a red blood cell, the cell shrinks, or crenates. This has the effect of concentrating the solutes left in the cell, making the cytosol denser and interfering with diffusion within the cell. The cell's ability to function will be compromised and may also result in the death of the cell.

Various living things have ways of controlling the effects of osmosis—a mechanism called osmoregulation. Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis in a hypotonic solution. The plasma membrane can only expand to the limit of the cell wall, so the cell will not

lyse. In fact, the cytoplasm in plants is always slightly hypertonic to the cellular environment, and water will always enter a cell if water is available. This inflow of water produces turgor pressure, which stiffens the cell walls of the plant ([\[link\]](#)). In nonwoody plants, turgor pressure supports the plant. Conversely, if the plant is not watered, the extracellular fluid will become hypertonic, causing water to leave the cell. In this condition, the cell does not shrink because the cell wall is not flexible. However, the cell membrane detaches from the wall and constricts the cytoplasm. This is called **plasmolysis**. Plants lose turgor pressure in this condition and wilt ([\[link\]](#)).



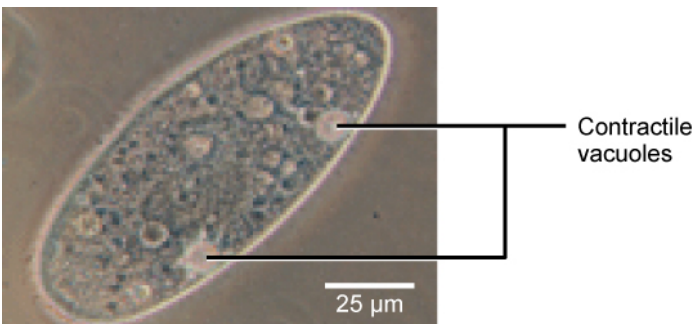
The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in.

(credit: modification of work by Mariana Ruiz Villareal)



Without adequate water, the plant on the left has lost turgor pressure, visible in its wilting; the turgor pressure is restored by watering it (right). (credit: Victor M. Vicente Selvas)

Tonicity is a concern for all living things. For example, paramecia and amoebas, which are protists that lack cell walls, have contractile vacuoles. This vesicle collects excess water from the cell and pumps it out, keeping the cell from lysing as it takes on water from its environment ([link](#)).



A paramecium's contractile vacuole, here visualized using bright field light microscopy at 480x magnification,

continuously pumps water out of the organism's body to keep it from bursting in a hypotonic medium.
(credit: modification of work by NIH;
scale-bar data from Matt Russell)

Many marine invertebrates have internal salt levels matched to their environments, making them isotonic with the water in which they live. Fish, however, must spend approximately five percent of their metabolic energy maintaining osmotic homeostasis. Freshwater fish live in an environment that is hypotonic to their cells. These fish actively take in salt through their gills and excrete diluted urine to rid themselves of excess water. Saltwater fish live in the reverse environment, which is hypertonic to their cells, and they secrete salt through their gills and excrete highly concentrated urine.

In vertebrates, the kidneys regulate the amount of water in the body. Osmoreceptors are specialized cells in the brain that monitor the concentration of solutes in the blood. If the levels of solutes increase beyond a certain range, a hormone is released that retards water loss through the kidney and dilutes the blood to safer levels. Animals also have high concentrations of albumin, which is produced by the liver, in their blood. This protein is too large to pass easily through plasma membranes and is a major factor in controlling the osmotic pressures applied to tissues.

Section Summary

The passive forms of transport, diffusion and osmosis, move materials of small molecular weight across membranes. Substances diffuse from areas of high concentration to areas of lower concentration, and this process continues until the substance is evenly distributed in a system. In solutions containing more than one substance, each type of molecule diffuses according to its own concentration gradient, independent of the diffusion of other substances. Many factors can affect the rate of diffusion, including concentration gradient, size of the particles that are diffusing, temperature of the system, and so on.

In living systems, diffusion of substances into and out of cells is mediated by the plasma membrane. Some materials diffuse readily through the membrane, but others are hindered, and their passage is made possible by specialized proteins, such as channels and transporters. The chemistry of living things occurs in aqueous solutions, and balancing the concentrations of those solutions is an ongoing problem. In living systems, diffusion of some substances would be slow or difficult without membrane proteins that facilitate transport.

Art Connections

Exercise:

Problem:

[\[link\]](#) A doctor injects a patient with what the doctor thinks is an isotonic saline solution. The patient dies, and an autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

Solution:

[\[link\]](#) No, it must have been hypotonic as a hypotonic solution would cause water to enter the cells, thereby making them burst.

Review Questions

Exercise:

Problem: Water moves via osmosis _____.

- a. throughout the cytoplasm
- b. from an area with a high concentration of other solutes to a lower one
- c. from an area with a high concentration of water to one of lower concentration

- d. from an area with a low concentration of water to one of higher concentration
-

Solution:

C

Exercise:

Problem:

The principal force driving movement in diffusion is the _____.

- a. temperature
 - b. particle size
 - c. concentration gradient
 - d. membrane surface area
-

Solution:

C

Exercise:

Problem: What problem is faced by organisms that live in fresh water?

- a. Their bodies tend to take in too much water.
 - b. They have no way of controlling their tonicity.
 - c. Only salt water poses problems for animals that live in it.
 - d. Their bodies tend to lose too much water to their environment.
-

Solution:

A

Free Response

Exercise:**Problem:**

Discuss why the following affect the rate of diffusion: molecular size, temperature, solution density, and the distance that must be traveled.

Solution:

Heavy molecules move more slowly than lighter ones. It takes more energy in the medium to move them along. Increasing or decreasing temperature increases or decreases the energy in the medium, affecting molecular movement. The denser a solution is, the harder it is for molecules to move through it, causing diffusion to slow down due to friction. Living cells require a steady supply of nutrients and a steady rate of waste removal. If the distance these substances need to travel is too great, diffusion cannot move nutrients and waste materials efficiently to sustain life.

Exercise:

Problem: Why does water move through a membrane?

Solution:

Water moves through a membrane in osmosis because there is a concentration gradient across the membrane of solute and solvent. The solute cannot effectively move to balance the concentration on both sides of the membrane, so water moves to achieve this balance.

Exercise:**Problem:**

Both of the regular intravenous solutions administered in medicine, normal saline and lactated Ringer's solution, are isotonic. Why is this important?

Solution:

Injection of isotonic solutions ensures that there will be no perturbation of the osmotic balance, and no water taken from tissues or added to them from the blood.

Glossary

aquaporin

channel protein that allows water through the membrane at a very high rate

carrier protein

membrane protein that moves a substance across the plasma membrane by changing its own shape

channel protein

membrane protein that allows a substance to pass through its hollow core across the plasma membrane

concentration gradient

area of high concentration adjacent to an area of low concentration

diffusion

passive process of transport of low-molecular weight material according to its concentration gradient

facilitated transport

process by which material moves down a concentration gradient (from high to low concentration) using integral membrane proteins

hypertonic

situation in which extracellular fluid has a higher osmolarity than the fluid inside the cell, resulting in water moving out of the cell

hypotonic

situation in which extracellular fluid has a lower osmolarity than the fluid inside the cell, resulting in water moving into the cell

isotonic

situation in which the extracellular fluid has the same osmolarity as the fluid inside the cell, resulting in no net movement of water into or out of the cell

osmolarity

total amount of substances dissolved in a specific amount of solution

osmosis

transport of water through a semipermeable membrane according to the concentration gradient of water across the membrane that results from the presence of solute that cannot pass through the membrane

passive transport

method of transporting material through a membrane that does not require energy

plasmolysis

detaching of the cell membrane from the cell wall and constriction of the cell membrane when a plant cell is in a hypertonic solution

selectively permeable

characteristic of a membrane that allows some substances through but not others

solute

substance dissolved in a liquid to form a solution

tonicity

amount of solute in a solution

transport protein

membrane protein that facilitates passage of a substance across a membrane by binding it

Active Transport

By the end of this section, you will be able to:

- Understand how electrochemical gradients affect ions
- Distinguish between primary active transport and secondary active transport

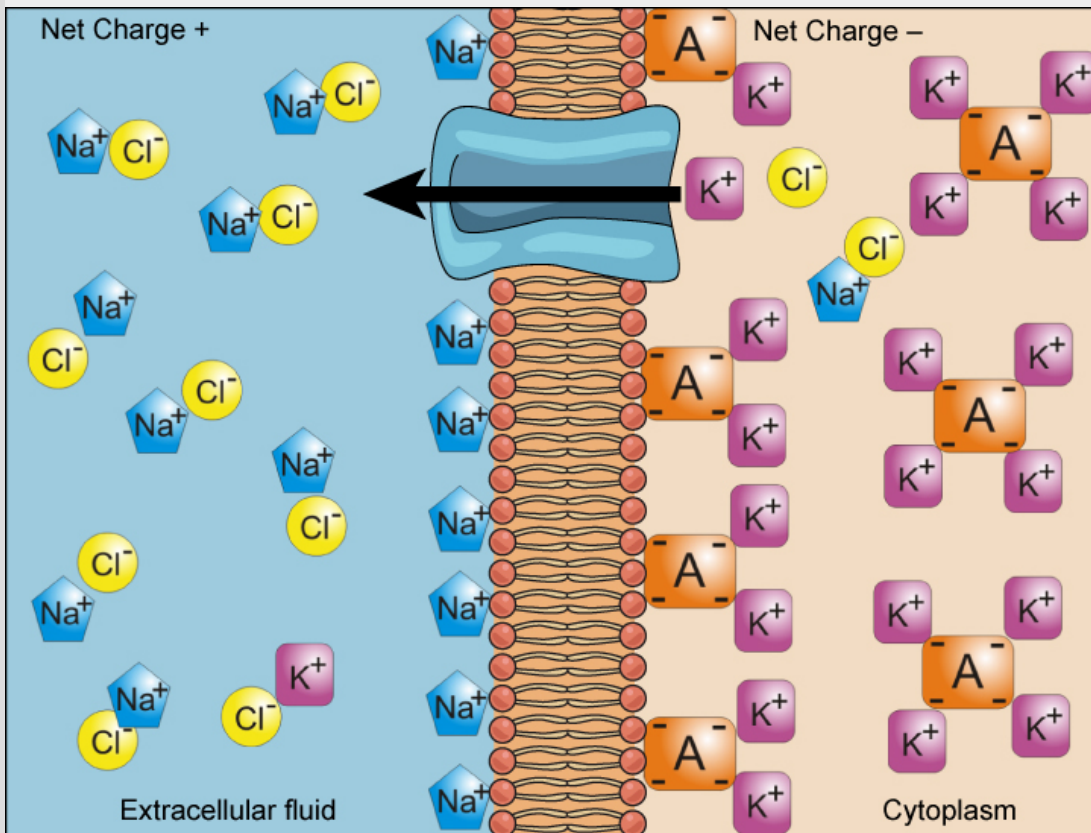
Active transport mechanisms require the use of the cell's energy, usually in the form of adenosine triphosphate (ATP). If a substance must move into the cell against its concentration gradient—that is, if the concentration of the substance inside the cell is greater than its concentration in the extracellular fluid (and vice versa)—the cell must use energy to move the substance. Some active transport mechanisms move small-molecular weight materials, such as ions, through the membrane. Other mechanisms transport much larger molecules.

Electrochemical Gradient

We have discussed simple concentration gradients—differential concentrations of a substance across a space or a membrane—but in living systems, gradients are more complex. Because ions move into and out of cells and because cells contain proteins that do not move across the membrane and are mostly negatively charged, there is also an electrical gradient, a difference of charge, across the plasma membrane. The interior of living cells is electrically negative with respect to the extracellular fluid in which they are bathed, and at the same time, cells have higher concentrations of potassium (K^+) and lower concentrations of sodium (Na^+) than does the extracellular fluid. So in a living cell, the concentration gradient of Na^+ tends to drive it into the cell, and the electrical gradient of Na^+ (a positive ion) also tends to drive it inward to the negatively charged interior. The situation is more complex, however, for other elements such as potassium. The electrical gradient of K^+ , a positive ion, also tends to drive it into the cell, but the concentration gradient of K^+ tends to drive K^+ out of the cell ([\[link\]](#)). The combined gradient of concentration and electrical charge that affects an ion is called its **electrochemical gradient**.

Note:

Art Connection



Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. (credit: "Synaptitude"/Wikimedia Commons)

Injection of a potassium solution into a person's blood is lethal; this is used in capital punishment and euthanasia. Why do you think a potassium solution injection is lethal?

Moving Against a Gradient

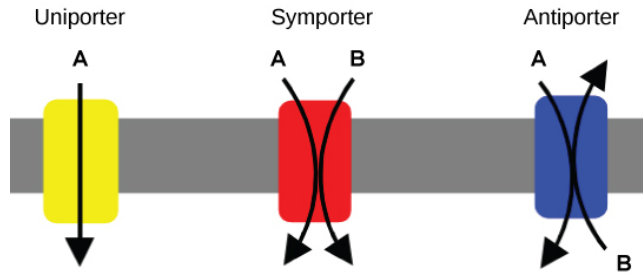
To move substances against a concentration or electrochemical gradient, the cell must use energy. This energy is harvested from ATP generated through the cell's metabolism. Active transport mechanisms, collectively called

pumps, work against electrochemical gradients. Small substances constantly pass through plasma membranes. Active transport maintains concentrations of ions and other substances needed by living cells in the face of these passive movements. Much of a cell's supply of metabolic energy may be spent maintaining these processes. (Most of a red blood cell's metabolic energy is used to maintain the imbalance between exterior and interior sodium and potassium levels required by the cell.) Because active transport mechanisms depend on a cell's metabolism for energy, they are sensitive to many metabolic poisons that interfere with the supply of ATP.

Two mechanisms exist for the transport of small-molecular weight material and small molecules. **Primary active transport** moves ions across a membrane and creates a difference in charge across that membrane, which is directly dependent on ATP. **Secondary active transport** describes the movement of material that is due to the electrochemical gradient established by primary active transport that does not directly require ATP.

Carrier Proteins for Active Transport

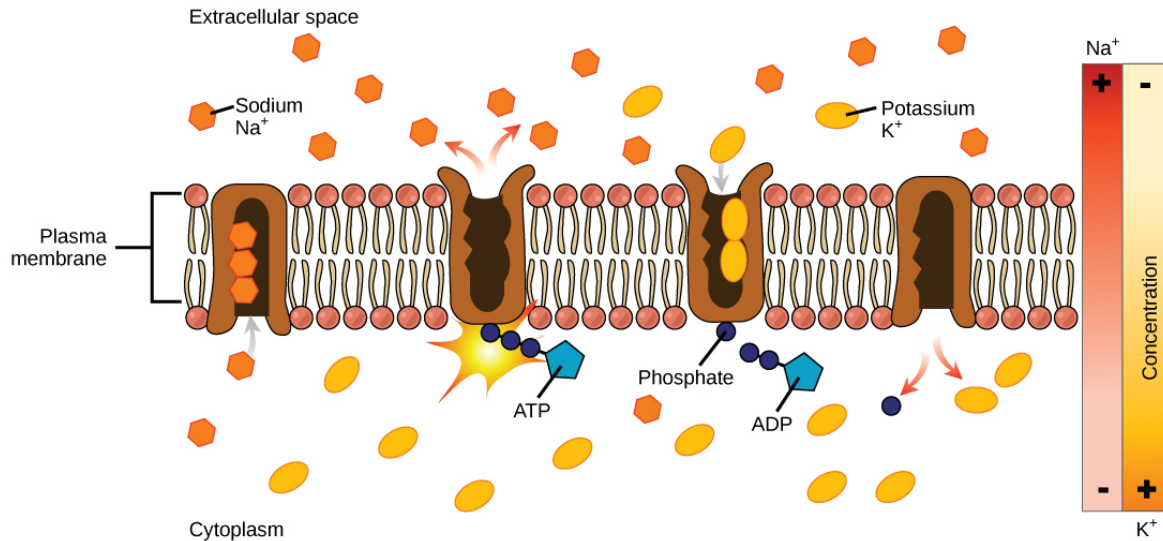
An important membrane adaption for active transport is the presence of specific carrier proteins or pumps to facilitate movement: there are three types of these proteins or **transporters** ([\[link\]](#)). A **uniporter** carries one specific ion or molecule. A **symporter** carries two different ions or molecules, both in the same direction. An **antiporter** also carries two different ions or molecules, but in different directions. All of these transporters can also transport small, uncharged organic molecules like glucose. These three types of carrier proteins are also found in facilitated diffusion, but they do not require ATP to work in that process. Some examples of pumps for active transport are $\text{Na}^+\text{-K}^+$ ATPase, which carries sodium and potassium ions, and $\text{H}^+\text{-K}^+$ ATPase, which carries hydrogen and potassium ions. Both of these are antiporter carrier proteins. Two other carrier proteins are Ca^{2+} ATPase and H^+ ATPase, which carry only calcium and only hydrogen ions, respectively. Both are pumps.



A uniporter carries one molecule or ion. A symporter carries two different molecules or ions, both in the same direction. An antiporter also carries two different molecules or ions, but in different directions. (credit: modification of work by “Lupask”/Wikimedia Commons)

Primary Active Transport

The primary active transport that functions with the active transport of sodium and potassium allows secondary active transport to occur. The second transport method is still considered active because it depends on the use of energy as does primary transport ([link](#)).



Primary active transport moves ions across a membrane, creating an electrochemical gradient (electrogenic transport). (credit: modification of work by Mariana Ruiz Villareal)

One of the most important pumps in animal cells is the sodium-potassium pump (Na⁺-K⁺ ATPase), which maintains the electrochemical gradient (and the correct concentrations of Na⁺ and K⁺) in living cells. The sodium-potassium pump moves K⁺ into the cell while moving Na⁺ out at the same time, at a ratio of three Na⁺ for every two K⁺ ions moved in. The Na⁺-K⁺ ATPase exists in two forms, depending on its orientation to the interior or exterior of the cell and its affinity for either sodium or potassium ions. The process consists of the following six steps.

1. With the enzyme oriented towards the interior of the cell, the carrier has a high affinity for sodium ions. Three ions bind to the protein.
2. ATP is hydrolyzed by the protein carrier and a low-energy phosphate group attaches to it.
3. As a result, the carrier changes shape and re-orientates itself towards the exterior of the membrane. The protein's affinity for sodium decreases and the three sodium ions leave the carrier.
4. The shape change increases the carrier's affinity for potassium ions, and two such ions attach to the protein. Subsequently, the low-energy

- phosphate group detaches from the carrier.
5. With the phosphate group removed and potassium ions attached, the carrier protein repositions itself towards the interior of the cell.
 6. The carrier protein, in its new configuration, has a decreased affinity for potassium, and the two ions are released into the cytoplasm. The protein now has a higher affinity for sodium ions, and the process starts again.

Several things have happened as a result of this process. At this point, there are more sodium ions outside of the cell than inside and more potassium ions inside than out. For every three ions of sodium that move out, two ions of potassium move in. This results in the interior being slightly more negative relative to the exterior. This difference in charge is important in creating the conditions necessary for the secondary process. The sodium-potassium pump is, therefore, an **electrogenic pump** (a pump that creates a charge imbalance), creating an electrical imbalance across the membrane and contributing to the membrane potential.

Note:

Link to Learning



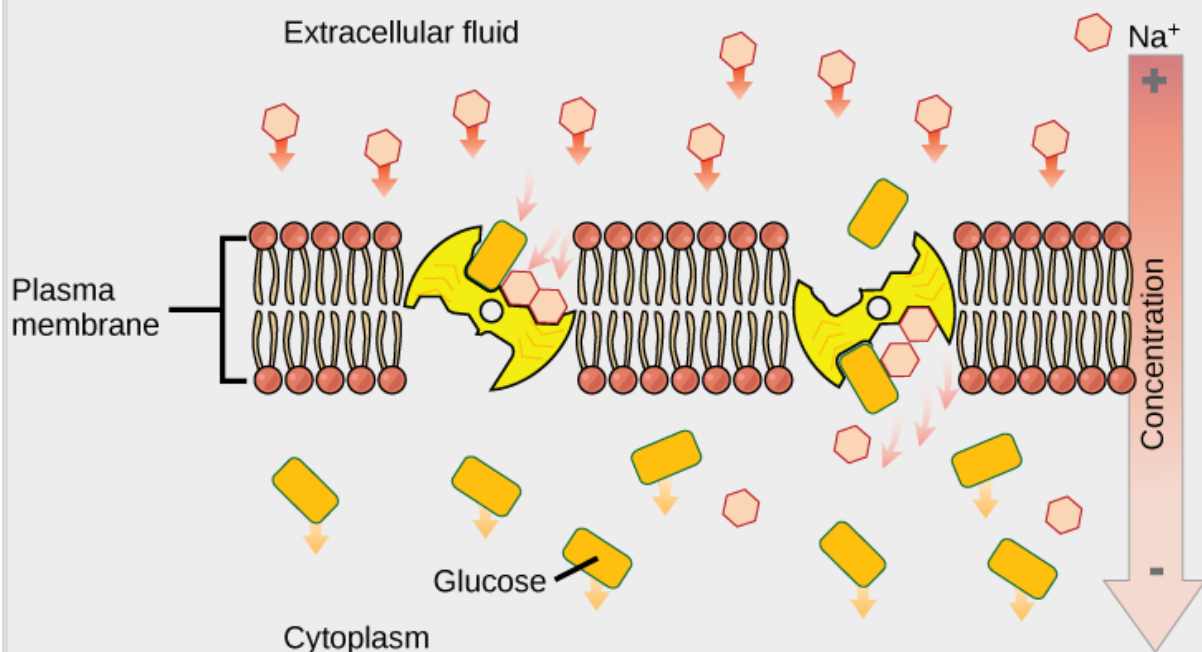
Watch this [video](#) to see a simulation of active transport in a sodium-potassium ATPase.

Secondary Active Transport (Co-transport)

Secondary active transport brings sodium ions, and possibly other compounds, into the cell. As sodium ion concentrations build outside of the plasma membrane because of the action of the primary active transport process, an electrochemical gradient is created. If a channel protein exists and is open, the sodium ions will be pulled through the membrane. This movement is used to transport other substances that can attach themselves to the transport protein through the membrane ([\[link\]](#)). Many amino acids, as well as glucose, enter a cell this way. This secondary process is also used to store high-energy hydrogen ions in the mitochondria of plant and animal cells for the production of ATP. The potential energy that accumulates in the stored hydrogen ions is translated into kinetic energy as the ions surge through the channel protein ATP synthase, and that energy is used to convert ADP into ATP.

Note:

Art Connection



An electrochemical gradient, created by primary active transport, can move other substances against their concentration gradients, a process called co-transport or secondary active transport. (credit: modification of work by Mariana Ruiz Villareal)

If the pH outside the cell decreases, would you expect the amount of amino acids transported into the cell to increase or decrease?

Section Summary

The combined gradient that affects an ion includes its concentration gradient and its electrical gradient. A positive ion, for example, might tend to diffuse into a new area, down its concentration gradient, but if it is diffusing into an area of net positive charge, its diffusion will be hampered by its electrical gradient. When dealing with ions in aqueous solutions, a combination of the electrochemical and concentration gradients, rather than just the concentration gradient alone, must be considered. Living cells need certain substances that exist inside the cell in concentrations greater than they exist in the extracellular space. Moving substances up their electrochemical gradients requires energy from the cell. Active transport uses energy stored in ATP to fuel this transport. Active transport of small molecular-sized materials uses integral proteins in the cell membrane to move the materials: These proteins are analogous to pumps. Some pumps, which carry out primary active transport, couple directly with ATP to drive their action. In co-transport (or secondary active transport), energy from primary transport can be used to move another substance into the cell and up its concentration gradient.

Art Connections

Exercise:

Problem:

[\[link\]](#) Injection of a potassium solution into a person's blood is lethal; this is used in capital punishment and euthanasia. Why do you think a potassium solution injection is lethal?

Solution:

[\[link\]](#) Cells typically have a high concentration of potassium in the cytoplasm and are bathed in a high concentration of sodium. Injection of potassium dissipates this electrochemical gradient. In heart muscle, the sodium/potassium potential is responsible for transmitting the signal that causes the muscle to contract. When this potential is dissipated, the signal can't be transmitted, and the heart stops beating. Potassium injections are also used to stop the heart from beating during surgery.

Exercise:

Problem:

[\[link\]](#) If the pH outside the cell decreases, would you expect the amount of amino acids transported into the cell to increase or decrease?

Solution:

[\[link\]](#) A decrease in pH means an increase in positively charged H^+ ions, and an increase in the electrical gradient across the membrane. The transport of amino acids into the cell will increase.

Review Questions

Exercise:

Problem:

Active transport must function continuously because _____.

- a. plasma membranes wear out
 - b. not all membranes are amphiphilic
 - c. facilitated transport opposes active transport
 - d. diffusion is constantly moving solutes in opposite directions
-

Solution:

D

Exercise:

Problem:

How does the sodium-potassium pump make the interior of the cell negatively charged?

- a. by expelling anions
 - b. by pulling in anions
 - c. by expelling more cations than are taken in
 - d. by taking in and expelling an equal number of cations
-

Solution:

C

Exercise:

Problem:

What is the combination of an electrical gradient and a concentration gradient called?

- a. potential gradient
 - b. electrical potential
 - c. concentration potential
 - d. electrochemical gradient
-

Solution:

D

Free Response

Exercise:

Problem:

Where does the cell get energy for active transport processes?

Solution:

The cell harvests energy from ATP produced by its own metabolism to power active transport processes, such as the activity of pumps.

Exercise:**Problem:**

How does the sodium-potassium pump contribute to the net negative charge of the interior of the cell?

Solution:

The sodium-potassium pump forces out three (positive) Na^+ ions for every two (positive) K^+ ions it pumps in, thus the cell loses a positive charge at every cycle of the pump.

Glossary

active transport

method of transporting material that requires energy

antiporter

transporter that carries two ions or small molecules in different directions

electrochemical gradient

gradient produced by the combined forces of an electrical gradient and a chemical gradient

electrogenic pump

pump that creates a charge imbalance

primary active transport

active transport that moves ions or small molecules across a membrane and may create a difference in charge across that membrane

pump

active transport mechanism that works against electrochemical gradients

secondary active transport

movement of material that is due to the electrochemical gradient established by primary active transport

symporter

transporter that carries two different ions or small molecules, both in the same direction

transporter

specific carrier proteins or pumps that facilitate movement

uniporter

transporter that carries one specific ion or molecule

Bulk Transport

By the end of this section, you will be able to:

- Describe endocytosis, including phagocytosis, pinocytosis, and receptor-mediated endocytosis
- Understand the process of exocytosis

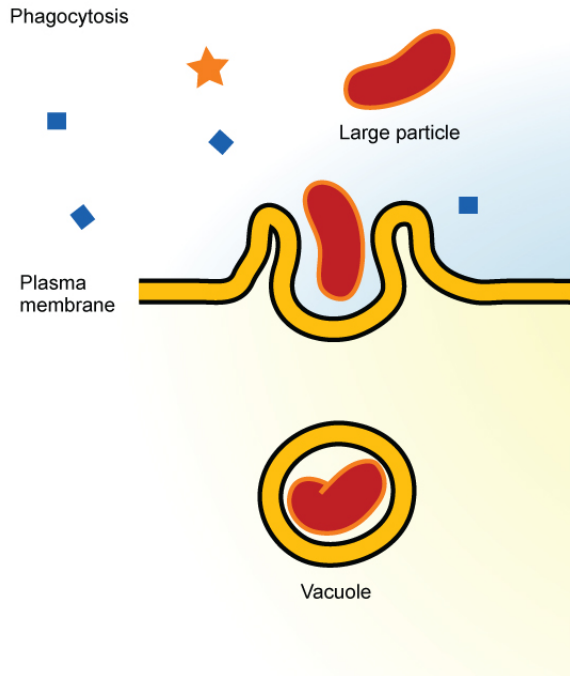
In addition to moving small ions and molecules through the membrane, cells also need to remove and take in larger molecules and particles (see [\[link\]](#) for examples). Some cells are even capable of engulfing entire unicellular microorganisms. You might have correctly hypothesized that the uptake and release of large particles by the cell requires energy. A large particle, however, cannot pass through the membrane, even with energy supplied by the cell.

Endocytosis

Endocytosis is a type of active transport that moves particles, such as large molecules, parts of cells, and even whole cells, into a cell. There are different variations of endocytosis, but all share a common characteristic: The plasma membrane of the cell invaginates, forming a pocket around the target particle. The pocket pinches off, resulting in the particle being contained in a newly created intracellular vesicle formed from the plasma membrane.

Phagocytosis

Phagocytosis (the condition of “cell eating”) is the process by which large particles, such as cells or relatively large particles, are taken in by a cell. For example, when microorganisms invade the human body, a type of white blood cell called a neutrophil will remove the invaders through this process, surrounding and engulfing the microorganism, which is then destroyed by the neutrophil ([\[link\]](#)).

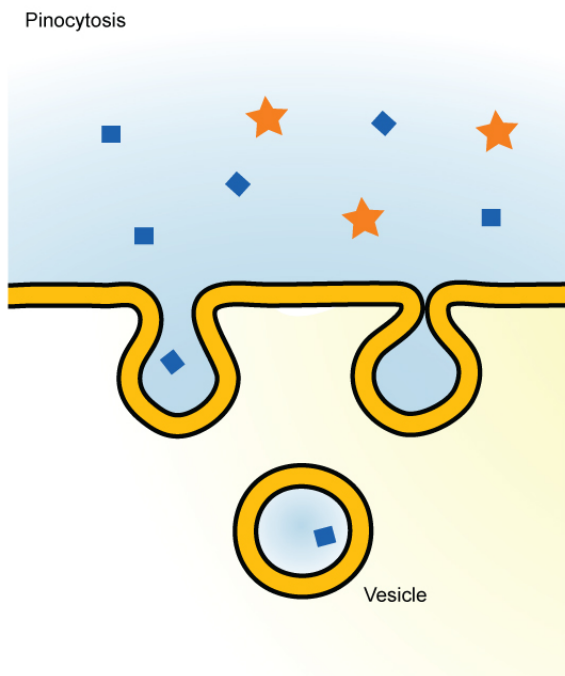


In phagocytosis, the cell membrane surrounds the particle and engulfs it. (credit: Mariana Ruiz Villareal)

In preparation for phagocytosis, a portion of the inward-facing surface of the plasma membrane becomes coated with a protein called **clathrin**, which stabilizes this section of the membrane. The coated portion of the membrane then extends from the body of the cell and surrounds the particle, eventually enclosing it. Once the vesicle containing the particle is enclosed within the cell, the clathrin disengages from the membrane and the vesicle merges with a lysosome for the breakdown of the material in the newly formed compartment (endosome). When accessible nutrients from the degradation of the vesicular contents have been extracted, the newly formed endosome merges with the plasma membrane and releases its contents into the extracellular fluid. The endosomal membrane again becomes part of the plasma membrane.

Pinocytosis

A variation of endocytosis is called **pinocytosis**. This literally means “cell drinking” and was named at a time when the assumption was that the cell was purposefully taking in extracellular fluid. In reality, this is a process that takes in molecules, including water, which the cell needs from the extracellular fluid. Pinocytosis results in a much smaller vesicle than does phagocytosis, and the vesicle does not need to merge with a lysosome ([\[link\]](#)).



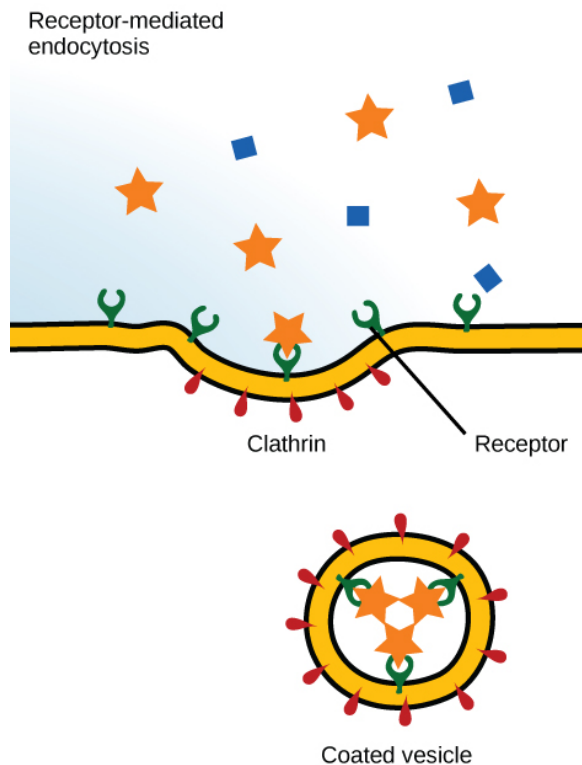
In pinocytosis, the cell membrane invaginates, surrounds a small volume of fluid, and pinches off. (credit: Mariana Ruiz Villareal)

A variation of pinocytosis is called **potocytosis**. This process uses a coating protein, called **caveolin**, on the cytoplasmic side of the plasma membrane,

which performs a similar function to clathrin. The cavities in the plasma membrane that form the vacuoles have membrane receptors and lipid rafts in addition to caveolin. The vacuoles or vesicles formed in caveolae (singular caveola) are smaller than those in pinocytosis. Potocytosis is used to bring small molecules into the cell and to transport these molecules through the cell for their release on the other side of the cell, a process called transcytosis.

Receptor-mediated Endocytosis

A targeted variation of endocytosis employs receptor proteins in the plasma membrane that have a specific binding affinity for certain substances ([link](#)).



In receptor-mediated endocytosis, uptake of

substances by the cell is targeted to a single type of substance that binds to the receptor on the external surface of the cell membrane.
(credit: modification of work by Mariana Ruiz Villareal)

In **receptor-mediated endocytosis**, as in phagocytosis, clathrin is attached to the cytoplasmic side of the plasma membrane. If uptake of a compound is dependent on receptor-mediated endocytosis and the process is ineffective, the material will not be removed from the tissue fluids or blood. Instead, it will stay in those fluids and increase in concentration. Some human diseases are caused by the failure of receptor-mediated endocytosis. For example, the form of cholesterol termed low-density lipoprotein or LDL (also referred to as “bad” cholesterol) is removed from the blood by receptor-mediated endocytosis. In the human genetic disease familial hypercholesterolemia, the LDL receptors are defective or missing entirely. People with this condition have life-threatening levels of cholesterol in their blood, because their cells cannot clear LDL particles from their blood.

Although receptor-mediated endocytosis is designed to bring specific substances that are normally found in the extracellular fluid into the cell, other substances may gain entry into the cell at the same site. Flu viruses, diphtheria, and cholera toxin all have sites that cross-react with normal receptor-binding sites and gain entry into cells.

Note:

[Link to Learning](#)

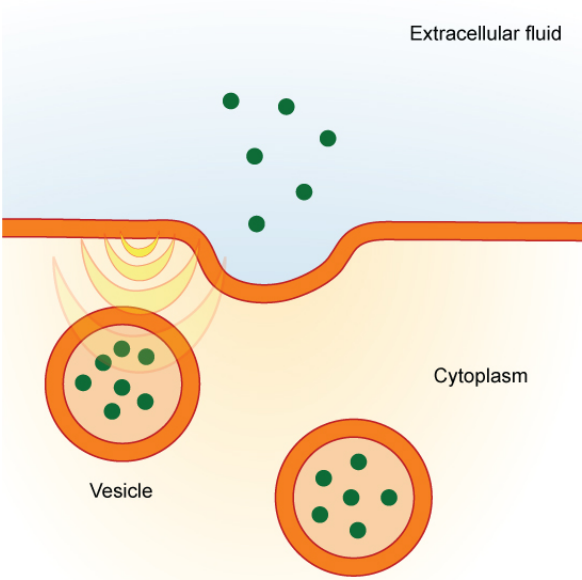


See receptor-mediated endocytosis in action, and click on different [parts](#) for a focused animation.

Exocytosis

The reverse process of moving material into a cell is the process of exocytosis. **Exocytosis** is the opposite of the processes discussed above in that its purpose is to expel material from the cell into the extracellular fluid. Waste material is enveloped in a membrane and fuses with the interior of the plasma membrane. This fusion opens the membranous envelope on the exterior of the cell, and the waste material is expelled into the extracellular space ([link](#)). Other examples of cells releasing molecules via exocytosis include the secretion of proteins of the extracellular matrix and secretion of neurotransmitters into the synaptic cleft by synaptic vesicles.

Exocytosis



In exocytosis, vesicles containing substances fuse with the plasma membrane. The contents are then released to the exterior of the cell.
(credit: modification of work by Mariana Ruiz Villareal)

Methods of Transport, Energy Requirements, and Types of Material Transported		
Transport Method	Active/Passive	Material Transported

Methods of Transport, Energy Requirements, and Types of Material Transported		
Transport Method	Active/Passive	Material Transported
Diffusion	Passive	Small-molecular weight material
Osmosis	Passive	Water
Facilitated transport/diffusion	Passive	Sodium, potassium, calcium, glucose
Primary active transport	Active	Sodium, potassium, calcium
Secondary active transport	Active	Amino acids, lactose
Phagocytosis	Active	Large macromolecules, whole cells, or cellular structures
Pinocytosis and potocytosis	Active	Small molecules (liquids/water)
Receptor-mediated endocytosis	Active	Large quantities of macromolecules

Section Summary

Active transport methods require the direct use of ATP to fuel the transport. Large particles, such as macromolecules, parts of cells, or whole cells, can be engulfed by other cells in a process called phagocytosis. In phagocytosis,

a portion of the membrane invaginates and flows around the particle, eventually pinching off and leaving the particle entirely enclosed by an envelope of plasma membrane. Vesicle contents are broken down by the cell, with the particles either used as food or dispatched. Pinocytosis is a similar process on a smaller scale. The plasma membrane invaginates and pinches off, producing a small envelope of fluid from outside the cell. Pinocytosis imports substances that the cell needs from the extracellular fluid. The cell expels waste in a similar but reverse manner: it pushes a membranous vacuole to the plasma membrane, allowing the vacuole to fuse with the membrane and incorporate itself into the membrane structure, releasing its contents to the exterior.

Review Questions

Exercise:

Problem:What happens to the membrane of a vesicle after exocytosis?

- a. It leaves the cell.
- b. It is disassembled by the cell.
- c. It fuses with and becomes part of the plasma membrane.
- d. It is used again in another exocytosis event.

Solution:

C

Exercise:

Problem:

Which transport mechanism can bring whole cells into a cell?

- a. pinocytosis
- b. phagocytosis
- c. facilitated transport
- d. primary active transport

Solution:

B

Exercise:**Problem:**

In what important way does receptor-mediated endocytosis differ from phagocytosis?

- a. It transports only small amounts of fluid.
- b. It does not involve the pinching off of membrane.
- c. It brings in only a specifically targeted substance.
- d. It brings substances into the cell, while phagocytosis removes substances.

Solution:

C

Free Response**Exercise:****Problem:**

Why is it important that there are different types of proteins in plasma membranes for the transport of materials into and out of a cell?

Solution:

The proteins allow a cell to select what compound will be transported, meeting the needs of the cell and not bringing in anything else.

Exercise:

Problem:

Why do ions have a difficult time getting through plasma membranes despite their small size?

Solution:

Ions are charged, and consequently, they are hydrophilic and cannot associate with the lipid portion of the membrane. Ions must be transported by carrier proteins or ion channels.

Glossary

caveolin

protein that coats the cytoplasmic side of the plasma membrane and participates in the process of liquid update by potocytosis

clathrin

protein that coats the inward-facing surface of the plasma membrane and assists in the formation of specialized structures, like coated pits, for phagocytosis

endocytosis

type of active transport that moves substances, including fluids and particles, into a cell

exocytosis

process of passing bulk material out of a cell

pinocytosis

a variation of endocytosis that imports macromolecules that the cell needs from the extracellular fluid

potocytosis

variation of pinocytosis that uses a different coating protein (caveolin) on the cytoplasmic side of the plasma membrane

receptor-mediated endocytosis

variation of endocytosis that involves the use of specific binding proteins in the plasma membrane for specific molecules or particles, and clathrin-coated pits that become clathrin-coated vesicles

Introduction

class="introduction"

A hummingbird needs energy to maintain prolonged periods of flight. The bird obtains its energy from taking in food and transforming the nutrients into energy through a series of biochemical reactions. The flight muscles in birds are extremely efficient in energy production. (credit: modification of work by Cory Zanker)



Virtually every task performed by living organisms requires energy. Energy is needed to perform heavy labor and exercise, but humans also use a great deal of energy while thinking, and even during sleep. In fact, the living cells of every organism constantly use energy. Nutrients and other molecules are imported, metabolized (broken down) and possibly synthesized into new molecules, modified if needed, transported around the cell, and may be distributed to the entire organism. For example, the large proteins that make up muscles are actively built from smaller molecules. Complex carbohydrates are broken down into simple sugars that the cell uses for energy. Just as energy is required to both build and demolish a building, energy is required for both the synthesis and breakdown of molecules. Additionally, signaling molecules such as hormones and neurotransmitters are transported between cells. Pathogenic bacteria and viruses are ingested and broken down by cells. Cells must also export waste and toxins to stay healthy, and many cells must swim or move surrounding materials via the beating motion of cellular appendages like cilia and flagella.

The cellular processes listed above require a steady supply of energy. From where, and in what form, does this energy come? How do living cells obtain energy, and how do they use it? This chapter will discuss different forms of energy and the physical laws that govern energy transfer. This chapter will

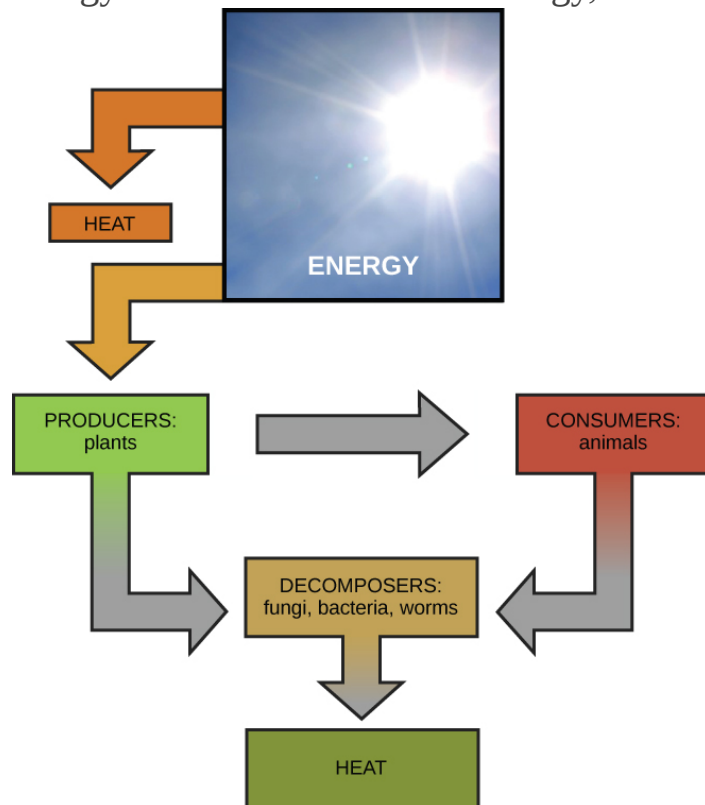
also describe how cells use energy and replenish it, and how chemical reactions in the cell are performed with great efficiency.

Energy and Metabolism

By the end of this section, you will be able to:

- Explain what metabolic pathways are and describe the two major types of metabolic pathways
- Discuss how chemical reactions play a role in energy transfer

Scientists use the term **bioenergetics** to discuss the concept of energy flow ([\[link\]](#)) through living systems, such as cells. Cellular processes such as the building and breaking down of complex molecules occur through stepwise chemical reactions. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Just as living things must continually consume food to replenish what has been used, cells must continually produce more energy to replenish that used by the many energy-requiring chemical reactions that constantly take place. All of the chemical reactions that take place inside cells, including those that use energy and those that release energy, are the cell's **metabolism**.



Most life forms on earth get their energy from the sun. Plants use

photosynthesis to capture sunlight, and herbivores eat those plants to obtain energy. Carnivores eat the herbivores, and decomposers digest plant and animal matter.

Metabolism of Carbohydrates

The metabolism of sugar (a simple carbohydrate) is a classic example of the many cellular processes that use and produce energy. Living things consume sugar as a major energy source, because sugar molecules have a great deal of energy stored within their bonds. The breakdown of glucose, a simple sugar, is described by the equation:

Equation:



Carbohydrates that are consumed have their origins in photosynthesizing organisms like plants ([link](#)). During photosynthesis, plants use the energy of sunlight to convert carbon dioxide gas (CO_2) into sugar molecules, like glucose ($\text{C}_6\text{H}_{12}\text{O}_6$). Because this process involves synthesizing a larger, energy-storing molecule, it requires an input of energy to proceed. The synthesis of glucose is described by this equation (notice that it is the reverse of the previous equation):

Equation:



During the chemical reactions of photosynthesis, energy is provided in the form of a very high-energy molecule called ATP, or adenosine triphosphate, which is the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work. The sugar (glucose) is stored as starch or

glycogen. Energy-storing polymers like these are broken down into glucose to supply molecules of ATP.

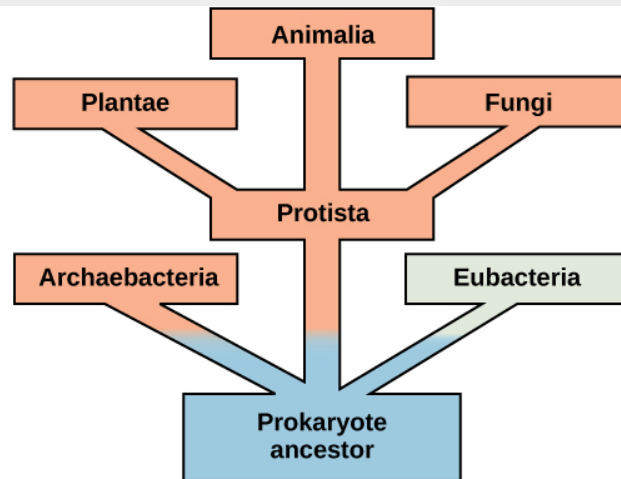
Solar energy is required to synthesize a molecule of glucose during the reactions of photosynthesis. In photosynthesis, light energy from the sun is initially transformed into chemical energy that is temporally stored in the energy carrier molecules ATP and NADPH (nicotinamide adenine dinucleotide phosphate). The stored energy in ATP and NADPH is then used later in photosynthesis to build one molecule of glucose from six molecules of CO_2 . This process is analogous to eating breakfast in the morning to acquire energy for your body that can be used later in the day. Under ideal conditions, energy from 18 molecules of ATP is required to synthesize one molecule of glucose during the reactions of photosynthesis. Glucose molecules can also be combined with and converted into other types of sugars. When sugars are consumed, molecules of glucose eventually make their way into each living cell of the organism. Inside the cell, each sugar molecule is broken down through a complex series of chemical reactions. The goal of these reactions is to harvest the energy stored inside the sugar molecules. The harvested energy is used to make high-energy ATP molecules, which can be used to perform work, powering many chemical reactions in the cell. The amount of energy needed to make one molecule of glucose from six molecules of carbon dioxide is 18 molecules of ATP and 12 molecules of NADPH (each one of which is energetically equivalent to three molecules of ATP), or a total of 54 molecule equivalents required for the synthesis of one molecule of glucose. This process is a fundamental and efficient way for cells to generate the molecular energy that they require.



Plants, like this oak tree and acorn, use energy from sunlight to make sugar and other organic molecules. Both plants and animals (like this squirrel) use cellular respiration to derive energy from the organic molecules originally produced by plants. (credit “acorn”: modification of work by Noel Reynolds; credit “squirrel”: modification of work by Dawn Huczek)

Metabolic Pathways

The processes of making and breaking down sugar molecules illustrate two types of metabolic pathways. A metabolic pathway is a series of interconnected biochemical reactions that convert a substrate molecule or molecules, step-by-step, through a series of metabolic intermediates, eventually yielding a final product or products. In the case of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as anabolic (building) and catabolic (breaking down) pathways, respectively. Consequently, metabolism is composed of building (anabolism) and degradation (catabolism).

Note:**Evolution Connection**

This tree shows the evolution of the various branches of life. The vertical dimension is time. Early life forms, in blue, used anaerobic metabolism to obtain energy from their surroundings.

Evolution of Metabolic Pathways

There is more to the complexity of metabolism than understanding the metabolic pathways alone. Metabolic complexity varies from organism to organism. Photosynthesis is the primary pathway in which photosynthetic organisms like plants (the majority of global synthesis is done by planktonic algae) harvest the sun's energy and convert it into carbohydrates. The by-product of photosynthesis is oxygen, required by some cells to carry out cellular respiration. During cellular respiration, oxygen aids in the catabolic breakdown of carbon compounds, like carbohydrates. Among the products of this catabolism are CO_2 and ATP. In addition, some eukaryotes perform catabolic processes without oxygen (fermentation); that is, they perform or use anaerobic metabolism. Organisms probably evolved anaerobic metabolism to survive (living organisms came into existence about 3.8 billion years ago, when the atmosphere lacked oxygen). Despite the differences between organisms and the complexity of metabolism, researchers have found that all branches of life share some of the same metabolic pathways, suggesting that all

organisms evolved from the same ancient common ancestor ([\[link\]](#)). Evidence indicates that over time, the pathways diverged, adding specialized enzymes to allow organisms to better adapt to their environment, thus increasing their chance to survive. However, the underlying principle remains that all organisms must harvest energy from their environment and convert it to ATP to carry out cellular functions.

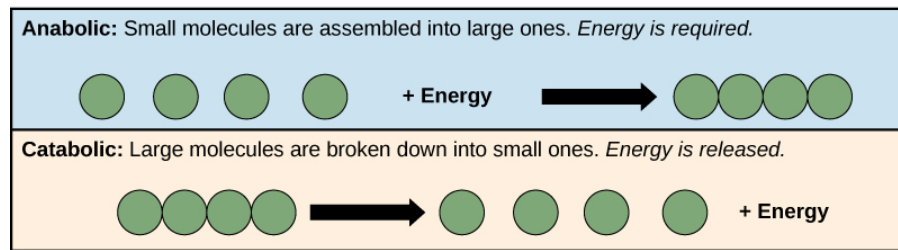
Anabolic and Catabolic Pathways

Anabolic pathways require an input of energy to synthesize complex molecules from simpler ones. Synthesizing sugar from CO_2 is one example. Other examples are the synthesis of large proteins from amino acid building blocks, and the synthesis of new DNA strands from nucleic acid building blocks. These biosynthetic processes are critical to the life of the cell, take place constantly, and demand energy provided by ATP and other high-energy molecules like NADH (nicotinamide adenine dinucleotide) and NADPH ([\[link\]](#)).

ATP is an important molecule for cells to have in sufficient supply at all times. The breakdown of sugars illustrates how a single molecule of glucose can store enough energy to make a great deal of ATP, 36 to 38 molecules. This is a **catabolic** pathway. Catabolic pathways involve the degradation (or breakdown) of complex molecules into simpler ones. Molecular energy stored in the bonds of complex molecules is released in catabolic pathways and harvested in such a way that it can be used to produce ATP. Other energy-storing molecules, such as fats, are also broken down through similar catabolic reactions to release energy and make ATP ([\[link\]](#)).

It is important to know that the chemical reactions of metabolic pathways don't take place spontaneously. Each reaction step is facilitated, or catalyzed, by a protein called an enzyme. Enzymes are important for catalyzing all types of biological reactions—those that require energy as well as those that release energy.

Metabolic pathways



Anabolic pathways are those that require energy to synthesize larger molecules. Catabolic pathways are those that generate energy by breaking down larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

Section Summary

Cells perform the functions of life through various chemical reactions. A cell's metabolism refers to the chemical reactions that take place within it. There are metabolic reactions that involve the breaking down of complex chemicals into simpler ones, such as the breakdown of large macromolecules. This process is referred to as catabolism, and such reactions are associated with a release of energy. On the other end of the spectrum, anabolism refers to metabolic processes that build complex molecules out of simpler ones, such as the synthesis of macromolecules. Anabolic processes require energy. Glucose synthesis and glucose breakdown are examples of anabolic and catabolic pathways, respectively.

Multiple Choice

Exercise:

Problem:

Energy is stored long-term in the bonds of _____ and used short-term to perform work from a(n) _____ molecule.

- a. ATP : glucose
 - b. an anabolic molecule : catabolic molecule
 - c. glucose : ATP
 - d. a catabolic molecule : anabolic molecule
-

Solution:

C

Exercise:

Problem:

DNA replication involves unwinding two strands of parent DNA, copying each strand to synthesize complementary strands, and releasing the parent and daughter DNA. Which of the following accurately describes this process?

- a. This is an anabolic process
 - b. This is a catabolic process
 - c. This is both anabolic and catabolic
 - d. This is a metabolic process but is neither anabolic nor catabolic
-

Solution:

A

Free Response

Exercise:

Problem:

Does physical exercise involve anabolic and/or catabolic processes?
Give evidence for your answer.

Solution:

Physical exercise involves both anabolic and catabolic processes. Body cells break down sugars to provide ATP to do the work necessary for exercise, such as muscle contractions. This is catabolism. Muscle cells also must repair muscle tissue damaged by exercise by building new muscle. This is anabolism.

Exercise:

Problem:

Name two different cellular functions that require energy that parallel human energy-requiring functions.

Solution:

Energy is required for cellular motion, through beating of cilia or flagella, as well as human motion, produced by muscle contraction. Cells also need energy to perform digestion, as humans require energy to digest food.

Glossary

anabolic

(also, anabolism) pathways that require an input of energy to synthesize complex molecules from simpler ones

bioenergetics

study of energy flowing through living systems

catabolic

(also, catabolism) pathways in which complex molecules are broken down into simpler ones

metabolism

all the chemical reactions that take place inside cells, including anabolism and catabolism

The Laws of Thermodynamics

By the end of this section, you will be able to:

- Discuss the concept of entropy
- Explain the first and second laws of thermodynamics

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter and its environment relevant to a particular case of energy transfer are classified as a system, and everything outside of that system is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. An open system is one in which energy can be transferred between the system and its surroundings. The stovetop system is open because heat can be lost into the air. A closed system is one that cannot transfer energy to its surroundings.

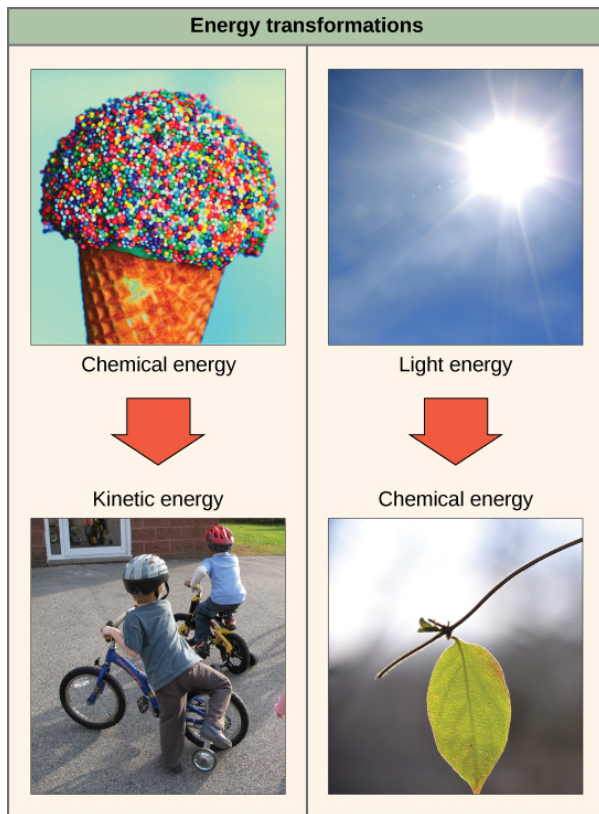
Biological organisms are open systems. Energy is exchanged between them and their surroundings, as they consume energy-storing molecules and release energy to the environment by doing work. Like all things in the physical world, energy is subject to the laws of physics. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

The First Law of Thermodynamics

The first law of thermodynamics deals with the total amount of energy in the universe. It states that this total amount of energy is constant. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight into the chemical energy

stored within organic molecules ([link](#)). Some examples of energy transformations are shown in [link](#).

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge very well. Chemical energy stored within organic molecules such as sugars and fats is transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the beating motion of cilia or flagella, contracting muscle fibers to create movement, and reproduction.



Shown are two examples of energy being transferred from one system to another and

transformed from one form to another. Humans can convert the chemical energy in food, like this ice cream cone, into kinetic energy (the energy of movement to ride a bicycle).

Plants can convert electromagnetic radiation (light energy) from the sun into chemical energy. (credit “ice cream”: modification of work by D. Sharon Pruitt; credit “kids on bikes”: modification of work by Michelle Rikken-Ransom; credit “leaf”: modification of work by Cory Zanker)

The Second Law of Thermodynamics

A living cell’s primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the second law of thermodynamics explains why these tasks are harder than they appear. None of the energy transfers we’ve discussed, along with all energy transfers and transformations in the universe, is completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, **heat energy** is defined as the energy transferred from one system to another that is not doing work. For example, when an airplane flies through the air, some of the energy of the flying plane is lost as heat energy due to friction with the surrounding air. This friction actually heats the air by temporarily increasing the speed of air molecules. Likewise, some energy is lost as heat energy during cellular metabolic reactions. This is good for warm-blooded creatures like us, because heat energy helps to maintain our body temperature. Strictly speaking, no energy transfer is completely efficient, because some energy is lost in an unusable form.

An important concept in physical systems is that of order and disorder (also known as randomness). The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of randomness or disorder within a system as **entropy**. High entropy means high disorder and low energy ([link](#)). To better understand entropy, think of a student's bedroom. If no energy or work were put into it, the room would quickly become messy. It would exist in a very disordered state, one of high entropy. Energy must be put into the system, in the form of the student doing work and putting everything away, in order to bring the room back to a state of cleanliness and order. This state is one of low entropy. Similarly, a car or house must be constantly maintained with work in order to keep it in an ordered state. Left alone, the entropy of the house or car gradually increases through rust and degradation. Molecules and chemical reactions have varying amounts of entropy as well. For example, as chemical reactions reach a state of equilibrium, entropy increases, and as molecules at a high concentration in one place diffuse and spread out, entropy also increases.

Note:

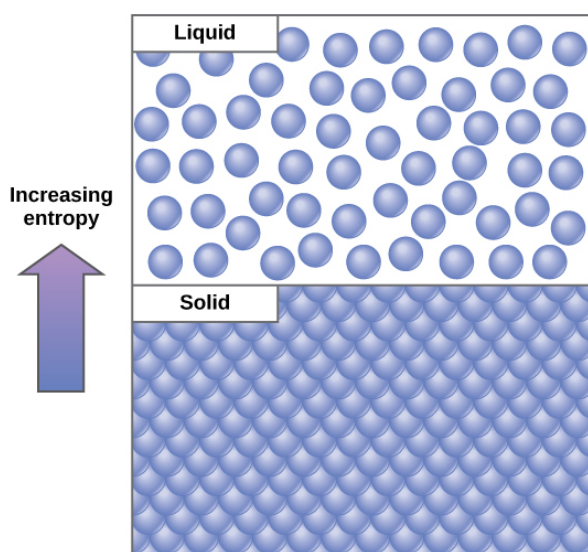
Scientific Connection

Transfer of Energy and the Resulting Entropy

Set up a simple experiment to understand how energy is transferred and how a change in entropy results.

1. Take a block of ice. This is water in solid form, so it has a high structural order. This means that the molecules cannot move very much and are in a fixed position. The temperature of the ice is 0°C . As a result, the entropy of the system is low.
2. Allow the ice to melt at room temperature. What is the state of molecules in the liquid water now? How did the energy transfer take place? Is the entropy of the system higher or lower? Why?
3. Heat the water to its boiling point. What happens to the entropy of the system when the water is heated?

All physical systems can be thought of in this way: Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy. As living systems take in energy-storing molecules and transform them through chemical reactions, they lose some amount of usable energy in the process, because no reaction is completely efficient. They also produce waste and by-products that aren't useful energy sources. This process increases the entropy of the system's surroundings. Since all energy transfers result in the loss of some usable energy, the second law of thermodynamics states that every energy transfer or transformation increases the entropy of the universe. Even though living things are highly ordered and maintain a state of low entropy, the entropy of the universe in total is constantly increasing due to the loss of usable energy with each energy transfer that occurs. Essentially, living things are in a continuous uphill battle against this constant increase in universal entropy.



Entropy is a measure of randomness or disorder in a system. Gases have higher entropy than liquids, and liquids have higher entropy than solids.

Section Summary

In studying energy, scientists use the term “system” to refer to the matter and its environment involved in energy transfers. Everything outside of the system is called the surroundings. Single cells are biological systems. Systems can be thought of as having a certain amount of order. It takes energy to make a system more ordered. The more ordered a system is, the lower its entropy. Entropy is a measure of the disorder of a system. As a system becomes more disordered, the lower its energy and the higher its entropy become.

A series of laws, called the laws of thermodynamics, describe the properties and processes of energy transfer. The first law states that the total amount of energy in the universe is constant. This means that energy can't be created or destroyed, only transferred or transformed. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy, resulting in a more disordered system. In other words, no energy transfer is completely efficient and tends toward disorder.

Review Questions

Exercise:

Problem:

Which of the following is not an example of an energy transformation?

- a. Turning on a light switch
- b. Solar panels at work
- c. Formation of static electricity
- d. None of the above

Solution:

A

Exercise:**Problem:**

Label each of the following systems as high or low entropy: i. the instant that a perfume bottle is sprayed compared with 30 seconds later, ii. an old 1950s car compared with a brand new car, and iii. a living cell compared with a dead cell.

- a. i. low, ii. high, iii. low
- b. i. low, ii. high, iii. high
- c. i. high, ii. low, iii. high
- d. i. high, ii. low, iii. Low

Solution:

A

Free Response**Exercise:****Problem:**

Imagine an elaborate ant farm with tunnels and passageways through the sand where ants live in a large community. Now imagine that an earthquake shook the ground and demolished the ant farm. In which of these two scenarios, before or after the earthquake, was the ant farm system in a state of higher or lower entropy?

Solution:

The ant farm had lower entropy before the earthquake because it was a highly ordered system. After the earthquake, the system became much more disordered and had higher entropy.

Exercise:

Problem:

Energy transfers take place constantly in everyday activities. Think of two scenarios: cooking on a stove and driving. Explain how the second law of thermodynamics applies to these two scenarios.

Solution:

While cooking, food is heating up on the stove, but not all of the heat goes to cooking the food, some of it is lost as heat energy to the surrounding air, increasing entropy. While driving, cars burn gasoline to run the engine and move the car. This reaction is not completely efficient, as some energy during this process is lost as heat energy, which is why the hood and the components underneath it heat up while the engine is turned on. The tires also heat up because of friction with the pavement, which is additional energy loss. This energy transfer, like all others, also increases entropy.

Glossary

entropy (S)

measure of randomness or disorder within a system

heat

energy energy transferred from one system to another that is not work (energy of the motion of molecules or particles)

thermodynamics

study of energy and energy transfer involving physical matter

Potential, Kinetic, Free, and Activation Energy

By the end of this section, you will be able to:

- Define “energy”
- Explain the difference between kinetic and potential energy
- Discuss the concepts of free energy and activation energy
- Describe endergonic and exergonic reactions

Energy is defined as the ability to do work. As you’ve learned, energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. While these are all familiar types of energy that one can see or feel, there is another type of energy that is much less tangible. This energy is associated with something as simple as an object held above the ground. In order to appreciate the way energy flows into and out of biological systems, it is important to understand more about the different types of energy that exist in the physical world.

Types of Energy

When an object is in motion, there is energy associated with that object. In the example of an airplane in flight, there is a great deal of energy associated with the motion of the airplane. This is because moving objects are capable of enacting a change, or doing work. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. However, a wrecking ball that is not in motion is incapable of performing work. Energy associated with objects in motion is called **kinetic energy**. A speeding bullet, a walking person, the rapid movement of molecules in the air (which produces heat), and electromagnetic radiation like light all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above a car with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The suspended wrecking ball has energy associated with it that is fundamentally different from the kinetic energy of objects in motion. This form of energy results from the fact that there is the *potential* for the wrecking ball to do work. If it is released, indeed it would do work. Because this type of energy refers to the potential

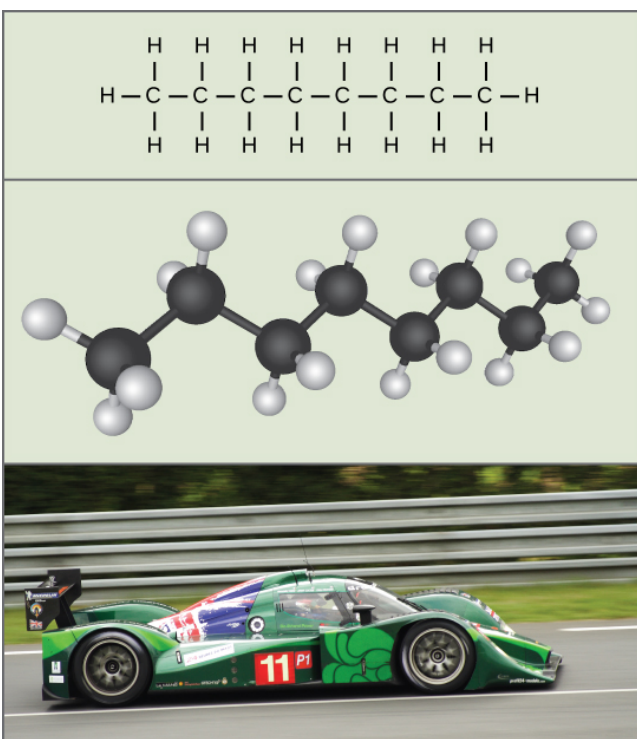
to do work, it is called **potential energy**. Objects transfer their energy between kinetic and potential in the following way: As the wrecking ball hangs motionless, it has 0 kinetic and 100 percent potential energy. Once it is released, its kinetic energy begins to increase because it builds speed due to gravity. At the same time, as it nears the ground, it loses potential energy. Somewhere mid-fall it has 50 percent kinetic and 50 percent potential energy. Just before it hits the ground, the ball has nearly lost its potential energy and has near-maximal kinetic energy. Other examples of potential energy include the energy of water held behind a dam ([link](#)), or a person about to skydive out of an airplane.



Water behind a dam has potential energy. Moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit “dam”: modification of work by "Pascal"/Flickr; credit “waterfall”: modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter (such as a child sitting on a tree branch), but also with the structure of matter. A spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. The very existence of living cells relies heavily on structural potential energy. On a chemical level, the bonds that hold the atoms of molecules together have potential energy. Remember that anabolic

cellular pathways require energy to synthesize complex molecules from simpler ones, and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called **chemical energy** ([link](#)). Chemical energy is responsible for providing living cells with energy from food. The release of energy is brought about by breaking the molecular bonds within fuel molecules.



The molecules in gasoline (octane, the chemical formula shown) contain chemical energy within the chemical bonds. This energy is transformed into kinetic energy

that allows a car to race on a racetrack. (credit “car”: modification of work by Russell Trow)

Note:

Link to Learning



Visit this [site](#) and select “A simple pendulum” on the menu (under “Harmonic Motion”) to see the shifting kinetic (K) and potential energy (U) of a pendulum in motion.

Free Energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is how is the energy associated with chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of **free energy** is used to quantitate these energy transfers. Free energy is called Gibbs free energy (abbreviated with the letter G) after Josiah Willard Gibbs, the scientist who developed the measurement. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat, resulting in entropy. Gibbs free energy specifically refers to the energy associated with a chemical reaction that is available after entropy is

accounted for. In other words, Gibbs free energy is usable energy, or energy that is available to do work.

Every chemical reaction involves a change in free energy, called delta G (ΔG). The change in free energy can be calculated for any system that undergoes such a change, such as a chemical reaction. To calculate ΔG , subtract the amount of energy lost to entropy (denoted as ΔS) from the total energy change of the system. This total energy change in the system is called **enthalpy** and is denoted as ΔH . The formula for calculating ΔG is as follows, where the symbol T refers to absolute temperature in Kelvin (degrees Celsius + 273):

Equation:

$$\Delta G = \Delta H - T\Delta S$$

The standard free energy change of a chemical reaction is expressed as an amount of energy per mole of the reaction product (either in kilojoules or kilocalories, kJ/mol or kcal/mol; 1 kJ = 0.239 kcal) under standard pH, temperature, and pressure conditions. Standard pH, temperature, and pressure conditions are generally calculated at pH 7.0 in biological systems, 25 degrees Celsius, and 100 kilopascals (1 atm pressure), respectively. It is important to note that cellular conditions vary considerably from these standard conditions, and so standard calculated ΔG values for biological reactions will be different inside the cell.

Endergonic Reactions and Exergonic Reactions

If energy is released during a chemical reaction, then the resulting value from the above equation will be a negative number. In other words, reactions that release energy have a $\Delta G < 0$. A negative ΔG also means that the products of the reaction have less free energy than the reactants, because they gave off some free energy during the reaction. Reactions that have a negative ΔG and consequently release free energy are called **exergonic reactions**. Think: exergonic means energy is exiting the system. These reactions are also referred to as spontaneous reactions, because they can

occur without the addition of energy into the system. Understanding which chemical reactions are spontaneous and release free energy is extremely useful for biologists, because these reactions can be harnessed to perform work inside the cell. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction that occurs immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction requires an input of energy rather than releasing energy, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions**, and they are non-spontaneous. An endergonic reaction will not take place on its own without the addition of free energy.

Let's revisit the example of the synthesis and breakdown of the food molecule, glucose. Remember that the building of complex molecules, such as sugars, from simpler ones is an anabolic process and requires energy. Therefore, the chemical reactions involved in anabolic processes are endergonic reactions. On the other hand, the catabolic process of breaking sugar down into simpler molecules releases energy in a series of exergonic reactions. Like the example of rust above, the breakdown of sugar involves spontaneous reactions, but these reactions don't occur instantaneously. [\[link\]](#) shows some other examples of endergonic and exergonic reactions. Later sections will provide more information about what else is required to make even spontaneous reactions happen more efficiently.

Note:

Art Connection



(a)



(b)



(c)



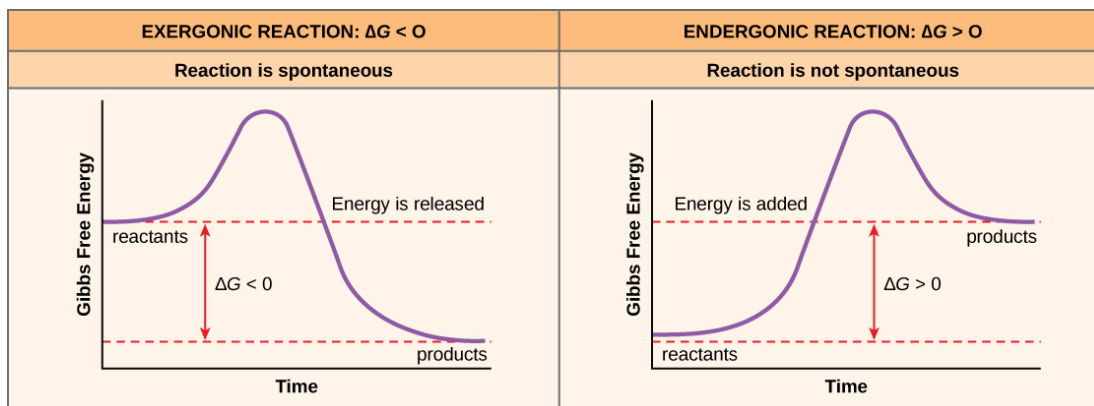
(d)

Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). These include (a) a compost pile decomposing, (b) a chick hatching from a fertilized egg, (c) sand art being destroyed, and (d) a ball rolling down a hill. (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by “Athlex”/Flickr; credit d: modification of work by Harry Malsch)

Look at each of the processes shown, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

An important concept in the study of metabolism and energy is that of chemical equilibrium. Most chemical reactions are reversible. They can

proceed in both directions, releasing energy into their environment in one direction, and absorbing it from the environment in the other direction ([link](#)). The same is true for the chemical reactions involved in cell metabolism, such as the breaking down and building up of proteins into and from individual amino acids, respectively. Reactants within a closed system will undergo chemical reactions in both directions until a state of equilibrium is reached. This state of equilibrium is one of the lowest possible free energy and a state of maximal entropy. Energy must be put into the system to push the reactants and products away from a state of equilibrium. Either reactants or products must be added, removed, or changed. If a cell were a closed system, its chemical reactions would reach equilibrium, and it would die because there would be insufficient free energy left to perform the work needed to maintain life. In a living cell, chemical reactions are constantly moving towards equilibrium, but never reach it. This is because a living cell is an open system. Materials pass in and out, the cell recycles the products of certain chemical reactions into other reactions, and chemical equilibrium is never reached. In this way, living organisms are in a constant energy-requiring, uphill battle against equilibrium and entropy. This constant supply of energy ultimately comes from sunlight, which is used to produce nutrients in the process of photosynthesis.



Exergonic and endergonic reactions result in changes in Gibbs free energy. Exergonic reactions release energy; endergonic reactions require energy to proceed.

Activation Energy

There is another important concept that must be considered regarding endergonic and exergonic reactions. Even exergonic reactions require a small amount of energy input to get going before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some energy in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy** (or free energy of activation) and is abbreviated E_A ([link](#)).

Why would an energy-releasing, negative ΔG reaction actually require some energy to proceed? The reason lies in the steps that take place during a chemical reaction. During chemical reactions, certain chemical bonds are broken and new ones are formed. For example, when a glucose molecule is broken down, bonds between the carbon atoms of the molecule are broken. Since these are energy-storing bonds, they release energy when broken. However, to get them into a state that allows the bonds to break, the molecule must be somewhat contorted. A small energy input is required to achieve this contorted state. This contorted state is called the **transition state**, and it is a high-energy, unstable state. For this reason, reactant molecules don't last long in their transition state, but very quickly proceed to the next steps of the chemical reaction. Free energy diagrams illustrate the energy profiles for a given reaction. Whether the reaction is exergonic or endergonic determines whether the products in the diagram will exist at a lower or higher energy state than both the reactants and the products. However, regardless of this measure, the transition state of the reaction exists at a higher energy state than the reactants, and thus, E_A is always positive.

Note:

Link to Learning



Watch an animation of the move from free energy to transition state at [this](#) site.

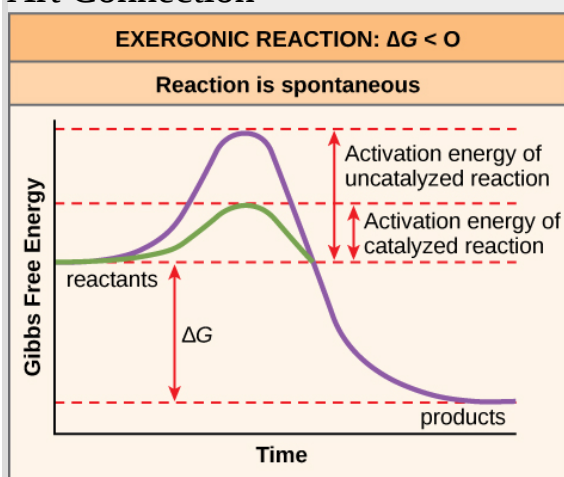
Where does the activation energy required by chemical reactants come from? The source of the activation energy needed to push reactions forward is typically heat energy from the surroundings. **Heat energy** (the total bond energy of reactants or products in a chemical reaction) speeds up the motion of molecules, increasing the frequency and force with which they collide; it also moves atoms and bonds within the molecule slightly, helping them reach their transition state. For this reason, heating up a system will cause chemical reactants within that system to react more frequently. Increasing the pressure on a system has the same effect. Once reactants have absorbed enough heat energy from their surroundings to reach the transition state, the reaction will proceed.

The activation energy of a particular reaction determines the rate at which it will proceed. The higher the activation energy, the slower the chemical reaction will be. The example of iron rusting illustrates an inherently slow reaction. This reaction occurs slowly over time because of its high E_A . Additionally, the burning of many fuels, which is strongly exergonic, will take place at a negligible rate unless their activation energy is overcome by sufficient heat from a spark. Once they begin to burn, however, the chemical reactions release enough heat to continue the burning process, supplying the activation energy for surrounding fuel molecules. Like these reactions outside of cells, the activation energy for most cellular reactions is too high for heat energy to overcome at efficient rates. In other words, in order for important cellular reactions to occur at appreciable rates (number of reactions per unit time), their activation energies must be lowered ([link](#)); this is referred to as catalysis. This is a very good thing as far as

living cells are concerned. Important macromolecules, such as proteins, DNA, and RNA, store considerable energy, and their breakdown is exergonic. If cellular temperatures alone provided enough heat energy for these exergonic reactions to overcome their activation barriers, the essential components of a cell would disintegrate.

Note:

Art Connection



Activation energy is the energy required for a reaction to proceed, and it is lower if the reaction is catalyzed. The horizontal axis of this diagram describes the sequence of events in time.

If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Section Summary

Energy comes in many different forms. Objects in motion do physical work, and kinetic energy is the energy of objects in motion. Objects that are not in motion may have the potential to do work, and thus, have potential energy. Molecules also have potential energy because the breaking of molecular bonds has the potential to release energy. Living cells depend on the harvesting of potential energy from molecular bonds to perform work. Free energy is a measure of energy that is available to do work. The free energy of a system changes during energy transfers such as chemical reactions, and this change is referred to as ΔG .

The ΔG of a reaction can be negative or positive, meaning that the reaction releases energy or consumes energy, respectively. A reaction with a negative ΔG that gives off energy is called an exergonic reaction. One with a positive ΔG that requires energy input is called an endergonic reaction. Exergonic reactions are said to be spontaneous, because their products have less energy than their reactants. The products of endergonic reactions have a higher energy state than the reactants, and so these are nonspontaneous reactions. However, all reactions (including spontaneous $-\Delta G$ reactions) require an initial input of energy in order to reach the transition state, at which they'll proceed. This initial input of energy is called the activation energy.

Art Connections

Exercise:

Problem:

[\[link\]](#) Look at each of the processes shown, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

Solution:

[\[link\]](#) A compost pile decomposing is an exergonic process; enthalpy increases (energy is released) and entropy increases (large molecules are broken down into smaller ones). A baby developing from a fertilized egg is an endergonic process; enthalpy decreases (energy is

absorbed) and entropy decreases. Sand art being destroyed is an exergonic process; there is no change in enthalpy, but entropy increases. A ball rolling downhill is an exergonic process; enthalpy decreases (energy is released), but there is no change in entropy.

Exercise:

Problem:

[\[link\]](#) If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Solution:

[\[link\]](#) No. We can store chemical energy because of the need to overcome the barrier to its breakdown.

Review Questions

Exercise:

Problem:

Consider a pendulum swinging. Which type(s) of energy is/are associated with the pendulum in the following instances: i. the moment at which it completes one cycle, just before it begins to fall back towards the other end, ii. the moment that it is in the middle between the two ends, iii. just before it reaches the end of one cycle (just before instant i.).

- a. i. potential and kinetic, ii. potential and kinetic, iii. kinetic
 - b. i. potential, ii. potential and kinetic, iii. potential and kinetic
 - c. i. potential, ii. kinetic, iii. potential and kinetic
 - d. i. potential and kinetic, ii. kinetic iii. kinetic
-

Solution:

C

Exercise:**Problem:**

Which of the following comparisons or contrasts between endergonic and exergonic reactions is false?

- a. Endergonic reactions have a positive ΔG and exergonic reactions have a negative ΔG
- b. Endergonic reactions consume energy and exergonic reactions release energy
- c. Both endergonic and exergonic reactions require a small amount of energy to overcome an activation barrier
- d. Endergonic reactions take place slowly and exergonic reactions take place quickly

Solution:

D

Exercise:**Problem:**

Which of the following is the best way to judge the relative activation energies between two given chemical reactions?

- a. Compare the ΔG values between the two reactions
- b. Compare their reaction rates
- c. Compare their ideal environmental conditions
- d. Compare the spontaneity between the two reactions

Solution:

B

Free Response

Exercise:**Problem:**

Explain in your own words the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:**Problem:**

Describe the position of the transition state on a vertical energy scale, from low to high, relative to the position of the reactants and products, for both endergonic and exergonic reactions.

Solution:

The transition state is always higher in energy than the reactants and the products of a reaction (therefore, above), regardless of whether the reaction is endergonic or exergonic.

Glossary

activation energy

energy necessary for reactions to occur

chemical energy

potential energy in chemical bonds that is released when those bonds are broken

endergonic

describes chemical reactions that require energy input

enthalpy

total energy of a system

exergonic

describes chemical reactions that release free energy

free energy

Gibbs free energy is the usable energy, or energy that is available to do work.

heat energy

total bond energy of reactants or products in a chemical reaction

kinetic energy

type of energy associated with objects or particles in motion

potential energy

type of energy that has the potential to do work; stored energy

transition state

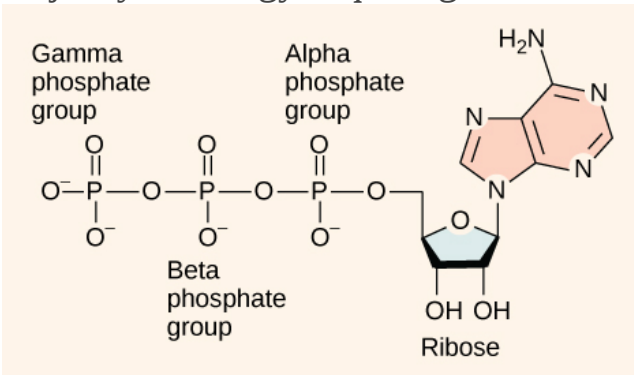
high-energy, unstable state (an intermediate form between the substrate and the product) occurring during a chemical reaction

ATP: Adenosine Triphosphate

By the end of this section, you will be able to:

- Explain the role of ATP as the cellular energy currency
- Describe how energy is released through hydrolysis of ATP

Even exergonic, energy-releasing reactions require a small amount of activation energy in order to proceed. However, consider endergonic reactions, which require much more energy input, because their products have more free energy than their reactants. Within the cell, where does energy to power such reactions come from? The answer lies with an energy-supplying molecule called **adenosine triphosphate**, or **ATP**. ATP is a small, relatively simple molecule ([\[link\]](#)), but within some of its bonds, it contains the potential for a quick burst of energy that can be harnessed to perform cellular work. This molecule can be thought of as the primary energy currency of cells in much the same way that money is the currency that people exchange for things they need. ATP is used to power the majority of energy-requiring cellular reactions.



ATP is the primary energy currency of the cell. It has an adenosine backbone with three phosphate groups attached.

As its name suggests, adenosine triphosphate is comprised of adenosine bound to three phosphate groups ([\[link\]](#)). Adenosine is a nucleoside consisting of the nitrogenous base adenine and a five-carbon sugar, ribose.

The three phosphate groups, in order of closest to furthest from the ribose sugar, are labeled alpha, beta, and gamma. Together, these chemical groups constitute an energy powerhouse. However, not all bonds within this molecule exist in a particularly high-energy state. Both bonds that link the phosphates are equally high-energy bonds (**phosphoanhydride bonds**) that, when broken, release sufficient energy to power a variety of cellular reactions and processes. These high-energy bonds are the bonds between the second and third (or beta and gamma) phosphate groups and between the first and second phosphate groups. The reason that these bonds are considered “high-energy” is because the products of such bond breaking—adenosine diphosphate (ADP) and one inorganic phosphate group (P_i)—have considerably lower free energy than the reactants: ATP and a water molecule. Because this reaction takes place with the use of a water molecule, it is considered a hydrolysis reaction. In other words, ATP is hydrolyzed into ADP in the following reaction:

Equation:



Like most chemical reactions, the hydrolysis of ATP to ADP is reversible. The reverse reaction regenerates ATP from $\text{ADP} + P_i$. Indeed, cells rely on the regeneration of ATP just as people rely on the regeneration of spent money through some sort of income. Since ATP hydrolysis releases energy, ATP regeneration must require an input of free energy. The formation of ATP is expressed in this equation:

Equation:



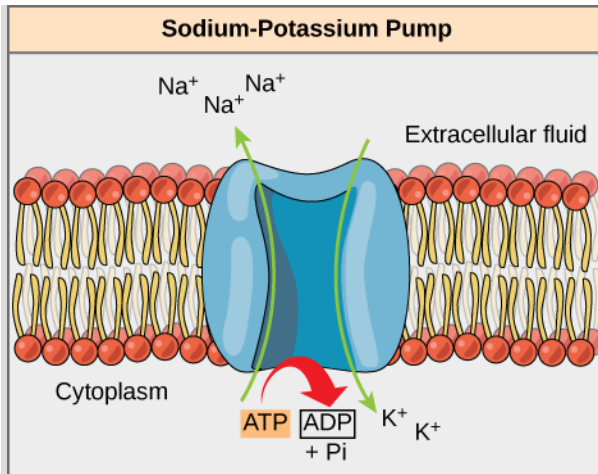
Two prominent questions remain with regard to the use of ATP as an energy source. Exactly how much free energy is released with the hydrolysis of ATP, and how is that free energy used to do cellular work? The calculated ΔG for the hydrolysis of one mole of ATP into ADP and P_i is -7.3 kcal/mole (-30.5 kJ/mol). Since this calculation is true under standard conditions, it would be expected that a different value exists under cellular conditions. In fact, the ΔG for the hydrolysis of one mole of ATP in a living

cell is almost double the value at standard conditions: -14 kcal/mol (-57 kJ/mol).

ATP is a highly unstable molecule. Unless quickly used to perform work, ATP spontaneously dissociates into $\text{ADP} + \text{P}_i$, and the free energy released during this process is lost as heat. The second question posed above, that is, how the energy released by ATP hydrolysis is used to perform work inside the cell, depends on a strategy called energy coupling. Cells couple the exergonic reaction of ATP hydrolysis with endergonic reactions, allowing them to proceed. One example of energy coupling using ATP involves a transmembrane ion pump that is extremely important for cellular function. This sodium-potassium pump (Na^+/K^+ pump) drives sodium out of the cell and potassium into the cell ([\[link\]](#)). A large percentage of a cell's ATP is spent powering this pump, because cellular processes bring a great deal of sodium into the cell and potassium out of the cell. The pump works constantly to stabilize cellular concentrations of sodium and potassium. In order for the pump to turn one cycle (exporting three Na^+ ions and importing two K^+ ions), one molecule of ATP must be hydrolyzed. When ATP is hydrolyzed, its gamma phosphate doesn't simply float away, but is actually transferred onto the pump protein. This process of a phosphate group binding to a molecule is called phosphorylation. As with most cases of ATP hydrolysis, a phosphate from ATP is transferred onto another molecule. In a phosphorylated state, the Na^+/K^+ pump has more free energy and is triggered to undergo a conformational change. This change allows it to release Na^+ to the outside of the cell. It then binds extracellular K^+ , which, through another conformational change, causes the phosphate to detach from the pump. This release of phosphate triggers the K^+ to be released to the inside of the cell. Essentially, the energy released from the hydrolysis of ATP is coupled with the energy required to power the pump and transport Na^+ and K^+ ions. ATP performs cellular work using this basic form of energy coupling through phosphorylation.

Note:

Art Connection



The sodium-potassium pump is an example of energy coupling.

The energy derived from exergonic ATP hydrolysis is used to pump sodium and potassium ions across the cell membrane.

The hydrolysis of one ATP molecule releases 7.3 kcal/mol of energy ($\Delta G = -7.3$ kcal/mol of energy). If it takes 2.1 kcal/mol of energy to move one Na^+ across the membrane ($\Delta G = +2.1$ kcal/mol of energy), how many sodium ions could be moved by the hydrolysis of one ATP molecule?

Often during cellular metabolic reactions, such as the synthesis and breakdown of nutrients, certain molecules must be altered slightly in their conformation to become substrates for the next step in the reaction series. One example is during the very first steps of cellular respiration, when a molecule of the sugar glucose is broken down in the process of glycolysis. In the first step of this process, ATP is required for the phosphorylation of glucose, creating a high-energy but unstable intermediate. This phosphorylation reaction powers a conformational change that allows the phosphorylated glucose molecule to be converted to the phosphorylated sugar fructose. Fructose is a necessary intermediate for glycolysis to move

forward. Here, the exergonic reaction of ATP hydrolysis is coupled with the endergonic reaction of converting glucose into a phosphorylated intermediate in the pathway. Once again, the energy released by breaking a phosphate bond within ATP was used for the phosphorylation of another molecule, creating an unstable intermediate and powering an important conformational change.

Note:

Link to Learning



See an interactive animation of the ATP-producing glycolysis process at this [site](#).

Section Summary

ATP is the primary energy-supplying molecule for living cells. ATP is made up of a nucleotide, a five-carbon sugar, and three phosphate groups. The bonds that connect the phosphates (phosphoanhydride bonds) have high-energy content. The energy released from the hydrolysis of ATP into ADP + P_i is used to perform cellular work. Cells use ATP to perform work by coupling the exergonic reaction of ATP hydrolysis with endergonic reactions. ATP donates its phosphate group to another molecule via a process known as phosphorylation. The phosphorylated molecule is at a higher-energy state and is less stable than its unphosphorylated form, and this added energy from the addition of the phosphate allows the molecule to undergo its endergonic reaction.

Art Connections

Exercise:

Problem:

[\[link\]](#) The hydrolysis of one ATP molecule releases 7.3 kcal/mol of energy ($\Delta G = -7.3$ kcal/mol of energy). If it takes 2.1 kcal/mol of energy to move one Na^+ across the membrane ($\Delta G = +2.1$ kcal/mol of energy), how many sodium ions could be moved by the hydrolysis of one ATP molecule?

Solution:

[\[link\]](#) Three sodium ions could be moved by the hydrolysis of one ATP molecule. The ΔG of the coupled reaction must be negative. Movement of three sodium ions across the membrane will take 6.3 kcal of energy ($2.1 \text{ kcal} \times 3 \text{ Na}^+ \text{ ions} = 6.3 \text{ kcal}$). Hydrolysis of ATP provides 7.3 kcal of energy, more than enough to power this reaction. Movement of four sodium ions across the membrane, however, would require 8.4 kcal of energy, more than one ATP molecule can provide.

Review Questions

Exercise:

Problem: The energy released by the hydrolysis of ATP is

- a. primarily stored between the alpha and beta phosphates
 - b. equal to -57 kcal/mol
 - c. harnessed as heat energy by the cell to perform work
 - d. providing energy to coupled reactions
-

Solution:

D

Exercise:**Problem:**

Which of the following molecules is likely to have the most potential energy?

- a. sucrose
- b. ATP
- c. glucose
- d. ADP

Solution:

A

Free Response**Exercise:****Problem:**

Do you think that the E_A for ATP hydrolysis is relatively low or high? Explain your reasoning.

Solution:

The activation energy for hydrolysis is very low. Not only is ATP hydrolysis an exergonic process with a large $-\Delta G$, but ATP is also a very unstable molecule that rapidly breaks down into $ADP + P_i$ if not utilized quickly. This suggests a very low E_A since it hydrolyzes so quickly.

Glossary

ATP

adenosine triphosphate, the cell's energy currency

phosphoanhydride bond

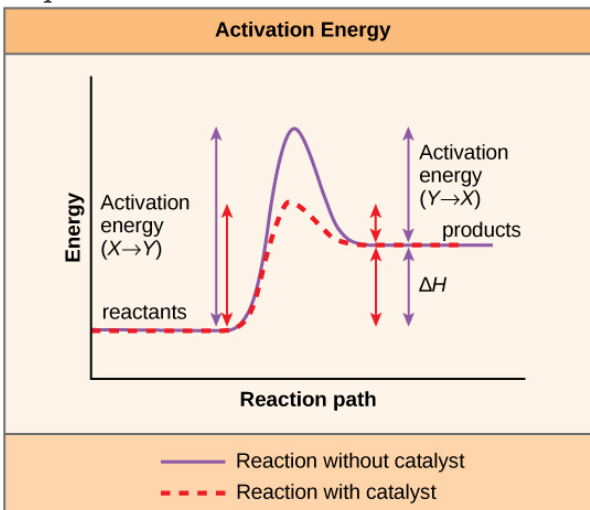
bond that connects phosphates in an ATP molecule

Enzymes

By the end of this section, you will be able to:

- Describe the role of enzymes in metabolic pathways
- Explain how enzymes function as molecular catalysts
- Discuss enzyme regulation by various factors

A substance that helps a chemical reaction to occur is a catalyst, and the special molecules that catalyze biochemical reactions are called enzymes. Almost all enzymes are proteins, made up of chains of amino acids, and they perform the critical task of lowering the activation energies of chemical reactions inside the cell. Enzymes do this by binding to the reactant molecules, and holding them in such a way as to make the chemical bond-breaking and bond-forming processes take place more readily. It is important to remember that enzymes don't change the ΔG of a reaction. In other words, they don't change whether a reaction is exergonic (spontaneous) or endergonic. This is because they don't change the free energy of the reactants or products. They only reduce the activation energy required to reach the transition state ([\[link\]](#)).



Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

Enzyme Active Site and Substrate Specificity

The chemical reactants to which an enzyme binds are the enzyme's **substrates**. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single-reactant substrate is broken down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction, both become modified, and leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme's **active site**. The active site is where the "action" happens, so to speak. Since enzymes are proteins, there is a unique combination of amino acid residues (also called side chains, or R groups) within the active site. Each residue is characterized by different properties. Residues can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of amino acid residues, their positions, sequences, structures, and properties, creates a very specific chemical environment within the active site. This specific environment is suited to bind, albeit briefly, to a specific chemical substrate (or substrates). Due to this jigsaw puzzle-like match between an enzyme and its substrates (which adapts to find the best fit between the transition state and the active site), enzymes are known for their specificity. The "best fit" results from the shape and the amino acid functional group's attraction to the substrate. There is a specifically matched enzyme for each substrate and, thus, for each chemical reaction; however, there is flexibility as well.

The fact that active sites are so perfectly suited to provide specific environmental conditions also means that they are subject to influences by the local environment. It is true that increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, increasing or decreasing the temperature outside of an optimal range can affect chemical bonds within the active site in such a way that they are less well suited to bind substrates. High temperatures will eventually cause enzymes, like other biological molecules, to **denature**, a process that changes the natural properties of a substance. Likewise, the pH of the local environment can also affect enzyme function. Active site amino acid residues have their own acidic or basic properties that are optimal for catalysis. These residues are sensitive to changes in pH that can impair the

way substrate molecules bind. Enzymes are suited to function best within a certain pH range, and, as with temperature, extreme pH values (acidic or basic) of the environment can cause enzymes to denature.

Induced Fit and Enzyme Function

For many years, scientists thought that enzyme-substrate binding took place in a simple “lock-and-key” fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a more refined view called **induced fit** ([link](#)). The induced-fit model expands upon the lock-and-key model by describing a more dynamic interaction between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme’s structure that confirms an ideal binding arrangement between the enzyme and the transition state of the substrate. This ideal binding maximizes the enzyme’s ability to catalyze its reaction.

Note:

Link to Learning

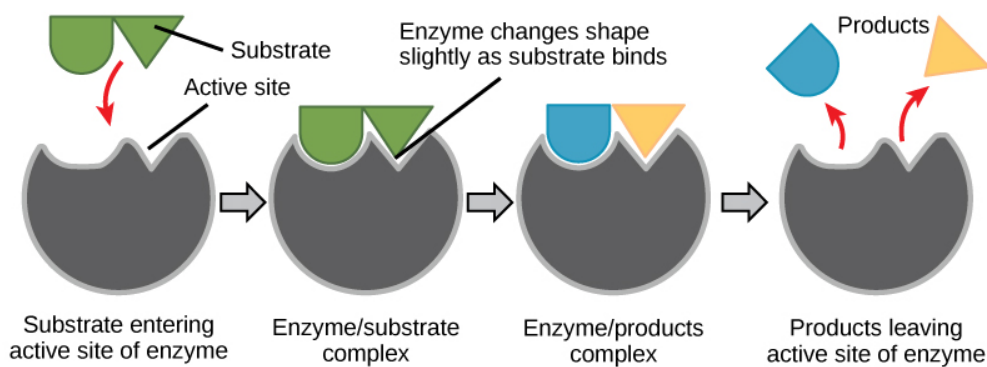


View an animation of induced fit at [this website](#).

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of many ways. On a basic level, enzymes promote chemical reactions that involve more than one substrate

by bringing the substrates together in an optimal orientation. The appropriate region (atoms and bonds) of one molecule is juxtaposed to the appropriate region of the other molecule with which it must react. Another way in which enzymes promote the reaction of their substrates is by creating an optimal environment within the active site for the reaction to occur. Certain chemical reactions might proceed best in a slightly acidic or non-polar environment. The chemical properties that emerge from the particular arrangement of amino acid residues within an active site create the perfect environment for an enzyme's specific substrates to react.

You've learned that the activation energy required for many reactions includes the energy involved in manipulating or slightly contorting chemical bonds so that they can easily break and allow others to reform. Enzymatic action can aid this process. The enzyme-substrate complex can lower the activation energy by contorting substrate molecules in such a way as to facilitate bond-breaking, helping to reach the transition state. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. The amino acid residues can provide certain ions or chemical groups that actually form covalent bonds with substrate molecules as a necessary step of the reaction process. In these cases, it is important to remember that the enzyme will always return to its original state at the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme is done catalyzing a reaction, it releases its product(s).



According to the induced-fit model, both enzyme and

substrate undergo dynamic conformational changes upon binding. The enzyme contorts the substrate into its transition state, thereby increasing the rate of the reaction.

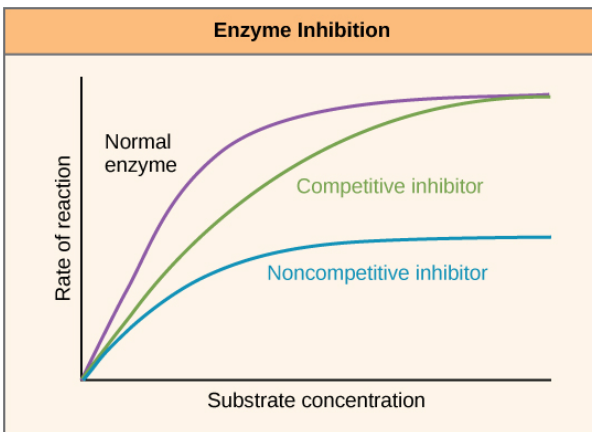
Control of Metabolism Through Enzyme Regulation

It would seem ideal to have a scenario in which all of the enzymes encoded in an organism's genome existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. In reality, this is far from the case. A variety of mechanisms ensure that this does not happen. Cellular needs and conditions vary from cell to cell, and change within individual cells over time. The required enzymes and energetic demands of stomach cells are different from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so do the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at which rates. This determination is tightly controlled. In certain cellular environments, enzyme activity is partly controlled by environmental factors, like pH and temperature. There are other mechanisms through which cells control the activity of enzymes and determine the rates at which various biochemical reactions will occur.

Regulation of Enzymes by Molecules

Enzymes can be regulated in ways that either promote or reduce their activity. There are many different kinds of molecules that inhibit or promote enzyme function, and various mechanisms exist for doing so. In some cases of enzyme inhibition, for example, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for active site binding ([\[link\]](#)). On the other hand, in noncompetitive inhibition, an inhibitor molecule binds to the enzyme in a location other than an allosteric site and still manages to block substrate binding to the active site.

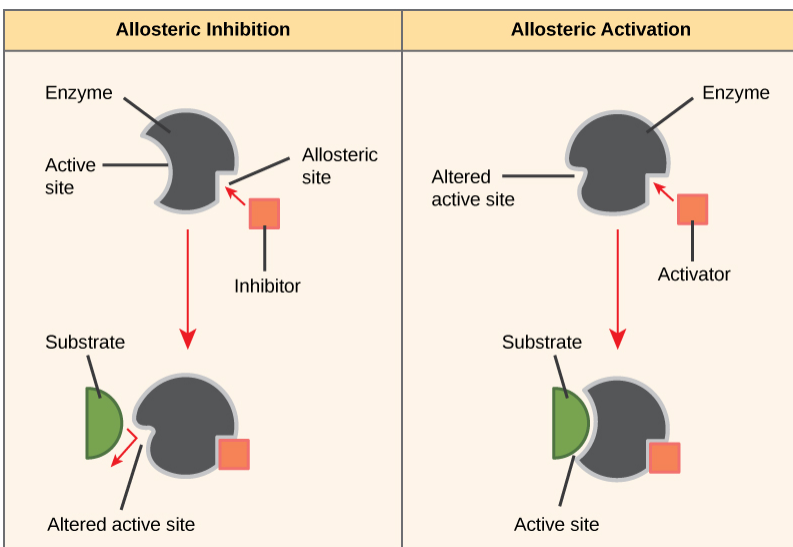


Competitive and noncompetitive inhibition affect the rate of reaction differently.

Competitive inhibitors affect the initial rate but do not affect the maximal rate, whereas noncompetitive inhibitors affect the maximal rate.

Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for

its substrate. This type of inhibition is called **allosteric inhibition** ([\[link\]](#)). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s).



Allosteric inhibitors modify the active site of the enzyme so that substrate binding is reduced or prevented. In contrast, allosteric activators modify the active site of the enzyme so that the affinity for the substrate increases.

Note:
Everyday Connection



Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Drug Discovery by Looking for Inhibitors of Key Enzymes in Specific Pathways

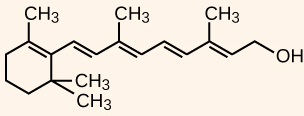
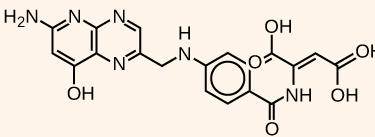
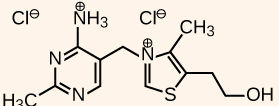
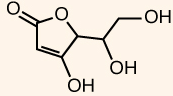
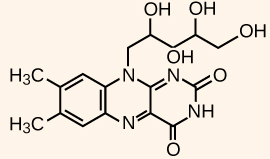
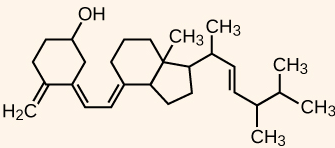
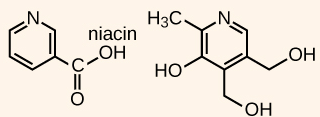
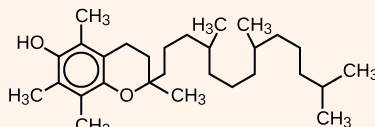
Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated is a key principle behind the development of many of the pharmaceutical drugs ([\[link\]](#)) on the market today. Biologists working in this field collaborate with other scientists, usually chemists, to design drugs.

Consider statins for example—which is the name given to the class of drugs that reduces cholesterol levels. These compounds are essentially inhibitors of the enzyme HMG-CoA reductase. HMG-CoA reductase is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the levels of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is effective in providing relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs developed? One of the first challenges in drug development is identifying the specific molecule that the drug is intended to target. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the

target alone is not sufficient; scientists also need to know how the target acts inside the cell and which reactions go awry in the case of disease. Once the target and the pathway are identified, then the actual process of drug design begins. During this stage, chemists and biologists work together to design and synthesize molecules that can either block or activate a particular reaction. However, this is only the beginning: both if and when a drug prototype is successful in performing its function, then it must undergo many tests from in vitro experiments to clinical trials before it can get FDA approval to be on the market.

Many enzymes don't work optimally, or even at all, unless bound to other specific non-protein helper molecules, either temporarily through ionic or hydrogen bonds or permanently through stronger covalent bonds. Two types of helper molecules are **cofactors** and **coenzymes**. Binding to these molecules promotes optimal conformation and function for their respective enzymes. Cofactors are inorganic ions such as iron (Fe^{++}) and magnesium (Mg^{++}). One example of an enzyme that requires a metal ion as a cofactor is the enzyme that builds DNA molecules, DNA polymerase, which requires bound zinc ion (Zn^{++}) to function. Coenzymes are organic helper molecules, with a basic atomic structure made up of carbon and hydrogen, which are required for enzyme action. The most common sources of coenzymes are dietary vitamins ([\[link\]](#)). Some vitamins are precursors to coenzymes and others act directly as coenzymes. Vitamin C is a coenzyme for multiple enzymes that take part in building the important connective tissue component, collagen. An important step in the breakdown of glucose to yield energy is catalysis by a multi-enzyme complex called pyruvate dehydrogenase. Pyruvate dehydrogenase is a complex of several enzymes that actually requires one cofactor (a magnesium ion) and five different organic coenzymes to catalyze its specific chemical reaction. Therefore, enzyme function is, in part, regulated by an abundance of various cofactors and coenzymes, which are supplied primarily by the diets of most organisms.

Dietary Vitamins	
Vitamin A 	Folic acid 
Vitamin B₁ 	Vitamin C 
Vitamin B₂ 	Vitamin D₂ (calciferol) 
Vitamin B₆ (pyridoxine) 	Vitamin E (alpha-tocopherol) 

Vitamins are important coenzymes or precursors of coenzymes, and are required for enzymes to function properly. Multivitamin capsules usually contain mixtures of all the vitamins at different percentages.

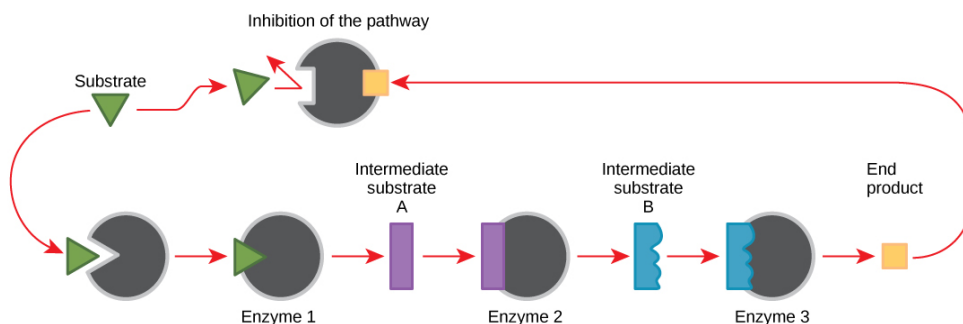
Enzyme Compartmentalization

In eukaryotic cells, molecules such as enzymes are usually compartmentalized into different organelles. This allows for yet another level of regulation of enzyme activity. Enzymes required only for certain cellular processes can be housed separately along with their substrates, allowing for more efficient chemical reactions. Examples of this sort of

enzyme regulation based on location and proximity include the enzymes involved in the latter stages of cellular respiration, which take place exclusively in the mitochondria, and the enzymes involved in the digestion of cellular debris and foreign materials, located within lysosomes.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. A major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, ions, and organic molecules, as you've learned. What other molecules in the cell provide enzymatic regulation, such as allosteric modulation, and competitive and noncompetitive inhibition? The answer is that a wide variety of molecules can perform these roles. Some of these molecules include pharmaceutical and non-pharmaceutical drugs, toxins, and poisons from the environment. Perhaps the most relevant sources of enzyme regulatory molecules, with respect to cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use the products of their own reactions for feedback inhibition of enzyme activity. **Feedback inhibition** involves the use of a reaction product to regulate its own further production ([\[link\]](#)). The cell responds to the abundance of specific products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.



Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream step, is an important regulatory mechanism in cells.

The production of both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that produces ATP. In this way, when ATP is abundant, the cell can prevent its further production. Remember that ATP is an unstable molecule that can spontaneously dissociate into ADP. If too much ATP were present in a cell, much of it would go to waste. On the other hand, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP through the catabolism of sugar.

Section Summary

Enzymes are chemical catalysts that accelerate chemical reactions at physiological temperatures by lowering their activation energy. Enzymes are usually proteins consisting of one or more polypeptide chains. Enzymes have an active site that provides a unique chemical environment, made up of certain amino acid R groups (residues). This unique environment is perfectly suited to convert particular chemical reactants for that enzyme, called substrates, into unstable intermediates called transition states. Enzymes and substrates are thought to bind with an induced fit, which means that enzymes undergo slight conformational adjustments upon substrate contact, leading to full, optimal binding. Enzymes bind to substrates and catalyze reactions in four different ways: bringing substrates together in an optimal orientation, compromising the bond structures of substrates so that bonds can be more easily broken, providing optimal environmental conditions for a reaction to occur, or participating directly in their chemical reaction by forming transient covalent bonds with the substrates.

Enzyme action must be regulated so that in a given cell at a given time, the desired reactions are being catalyzed and the undesired reactions are not. Enzymes are regulated by cellular conditions, such as temperature and pH. They are also regulated through their location within a cell, sometimes being compartmentalized so that they can only catalyze reactions under certain circumstances. Inhibition and activation of enzymes via other molecules are other important ways that enzymes are regulated. Inhibitors can act competitively, noncompetitively, or allosterically; noncompetitive inhibitors are usually allosteric. Activators can also enhance the function of enzymes allosterically. The most common method by which cells regulate the enzymes in metabolic pathways is through feedback inhibition. During feedback inhibition, the products of a metabolic pathway serve as inhibitors (usually allosteric) of one or more of the enzymes (usually the first committed enzyme of the pathway) involved in the pathway that produces them.

Review Questions

Exercise:

Problem: Which of the following is not true about enzymes:

- a. They increase ΔG of reactions
- b. They are usually made of amino acids
- c. They lower the activation energy of chemical reactions
- d. Each one is specific to the particular substrate(s) to which it binds

Solution:

A

Exercise:

Problem: An allosteric inhibitor does which of the following?

- a. Binds to an enzyme away from the active site and changes the conformation of the active site, increasing its affinity for substrate binding
- b. Binds to the active site and blocks it from binding substrate
- c. Binds to an enzyme away from the active site and changes the conformation of the active site, decreasing its affinity for the substrate
- d. Binds directly to the active site and mimics the substrate

Solution:

C

Exercise:

Problem:

Which of the following analogies best describe the induced-fit model of enzyme-substrate binding?

- a. A hug between two people
- b. A key fitting into a lock
- c. A square peg fitting through the square hole and a round peg fitting through the round hole of a children's toy
- d. The fitting together of two jigsaw puzzle pieces.

Solution:

A

Free Response

Exercise:

Problem:

With regard to enzymes, why are vitamins necessary for good health? Give examples.

Solution:

Most vitamins and minerals act as coenzymes and cofactors for enzyme action. Many enzymes require the binding of certain cofactors or coenzymes to be able to catalyze their reactions. Since enzymes catalyze many important reactions, it is critical to obtain sufficient vitamins and minerals from the diet and from supplements. Vitamin C (ascorbic acid) is a coenzyme necessary for the action of enzymes that build collagen, an important protein component of connective tissue throughout the body. Magnesium ion (Mg^{++}) is an important cofactor that is necessary for the enzyme pyruvate dehydrogenase to catalyze part of the pathway that breaks down sugar to produce energy. Vitamins cannot be produced in the human body and therefore must be obtained in the diet.

Exercise:**Problem:**

Explain in your own words how enzyme feedback inhibition benefits a cell.

Solution:

Feedback inhibition allows cells to control the amounts of metabolic products produced. If there is too much of a particular product relative to what the cell's needs, feedback inhibition effectively causes the cell to decrease production of that particular product. In general, this reduces the production of superfluous products and conserves energy, maximizing energy efficiency.

Glossary

active site

specific region of the enzyme to which the substrate binds

allosteric inhibition

inhibition by a binding event at a site different from the active site, which induces a conformational change and reduces the affinity of the enzyme for its substrate

coenzyme

small organic molecule, such as a vitamin or its derivative, which is required to enhance the activity of an enzyme

cofactor

inorganic ion, such as iron and magnesium ions, required for optimal regulation of enzyme activity

competitive inhibition

type of inhibition in which the inhibitor competes with the substrate molecule by binding to the active site of the enzyme

denature

process that changes the natural properties of a substance

feedback inhibition

effect of a product of a reaction sequence to decrease its further production by inhibiting the activity of the first enzyme in the pathway that produces it

induced fit

dynamic fit between the enzyme and its substrate, in which both components modify their structures to allow for ideal binding

substrate

molecule on which the enzyme acts

Introduction

class="introduction"

Each of us,
like these
other large
multicellula
r organisms,
begins life
as a
fertilized
egg. After
trillions of
cell
divisions,
each of us
develops
into a
complex,
multicellula
r organism.
(credit a:
modificatio
n of work
by Frank
Wouters;
credit b:
modificatio
n of work
by Ken
Cole,
USGS;
credit c:
modificatio
n of work
by Martin
Pettitt)



(a)

(b)

(c)

The ability to reproduce *in kind* is a basic characteristic of all living things. *In kind* means that the offspring of any organism closely resemble their parent or parents. Hippopotamuses give birth to hippopotamus calves, Joshua trees produce seeds from which Joshua tree seedlings emerge, and adult flamingos lay eggs that hatch into flamingo chicks. *In kind* does not generally mean *exactly the same*. Whereas many unicellular organisms and a few multicellular organisms can produce genetically identical clones of themselves through cell division, many single-celled organisms and most multicellular organisms reproduce regularly using another method. Sexual reproduction is the production by parents of two haploid cells and the fusion of two haploid cells to form a single, unique diploid cell. In most plants and animals, through tens of rounds of mitotic cell division, this diploid cell will develop into an adult organism. Haploid cells that are part of the sexual reproductive cycle are produced by a type of cell division called meiosis. Sexual reproduction, specifically meiosis and fertilization, introduces variation into offspring that may account for the evolutionary success of sexual reproduction. The vast majority of eukaryotic organisms, both multicellular and unicellular, can or must employ some form of meiosis and fertilization to reproduce.

The Process of Meiosis

By the end of this section, you will be able to:

- Describe the behavior of chromosomes during meiosis
- Describe cellular events during meiosis
- Explain the differences between meiosis and mitosis
- Explain the mechanisms within meiosis that generate genetic variation among the products of meiosis

Sexual reproduction requires **fertilization**, the union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. Haploid cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid. The number of sets of chromosomes in a cell is called its ploidy level. If the reproductive cycle is to continue, then the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division that reduces the number of chromosome sets.

Most animals and plants are diploid, containing two sets of chromosomes. In each **somatic cell** of the organism (all cells of a multicellular organism except the gametes or reproductive cells), the nucleus contains two copies of each chromosome, called homologous chromosomes. Somatic cells are sometimes referred to as “body” cells. Homologous chromosomes are matched pairs containing the same genes in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. Haploid cells, containing a single copy of each homologous chromosome, are found only within structures that give rise to either gametes or spores. **Spores** are haploid cells that can produce a haploid organism or can fuse with another spore to form a diploid cell. All animals and most plants produce eggs and sperm, or gametes. Some plants and all fungi produce spores.

The nuclear division that forms haploid cells, which is called **meiosis**, is related to mitosis. As you have learned, mitosis is the part of a cell

reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei are at the same ploidy level—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve this reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the major process and the stages are designated with a “I” or a “II.” Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. **Meiosis II**, in which the second round of meiotic division takes place, includes prophase II, prometaphase II, and so on.

Meiosis I

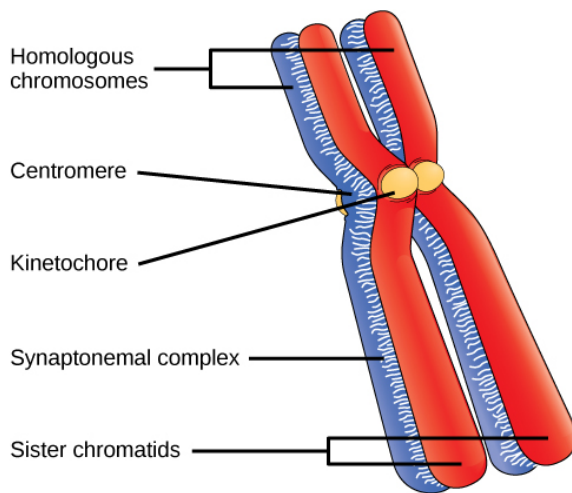
Meiosis is preceded by an interphase consisting of the G_1 , S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase, which is also called the first gap phase, is the first phase of the interphase and is focused on cell growth. The S phase is the second phase of interphase, during which the DNA of the chromosomes is replicated. Finally, the G_2 phase, also called the second gap phase, is the third and final phase of interphase; in this phase, the cell undergoes the final preparations for meiosis.

During DNA duplication in the S phase, each chromosome is replicated to produce two identical copies, called sister chromatids, that are held together at the centromere by **cohesin** proteins. Cohesin holds the chromatids together until anaphase II. The centrosomes, which are the structures that organize the microtubules of the meiotic spindle, also replicate. This prepares the cell to enter prophase I, the first meiotic phase.

Prophase I

Early in prophase I, before the chromosomes can be seen clearly microscopically, the homologous chromosomes are attached at their tips to the nuclear envelope by proteins. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. Recall that, in mitosis, homologous chromosomes do not pair together. In mitosis, homologous chromosomes line up end-to-end so that when they divide, each daughter cell receives a sister chromatid from both members of the homologous pair. The **synaptonemal complex**, a lattice of proteins between the homologous chromosomes, first forms at specific locations and then spreads to cover the entire length of the chromosomes. The tight pairing of the homologous chromosomes is called **synapsis**. In synapsis, the genes on the chromatids of the homologous chromosomes are aligned precisely with each other. The synaptonemal complex supports the exchange of chromosomal segments between non-sister homologous chromatids, a process called crossing over. Crossing over can be observed visually after the exchange as **chiasmata** (singular = chiasma) ([link](#)).

In species such as humans, even though the X and Y sex chromosomes are not homologous (most of their genes differ), they have a small region of homology that allows the X and Y chromosomes to pair up during prophase I. A partial synaptonemal complex develops only between the regions of homology.

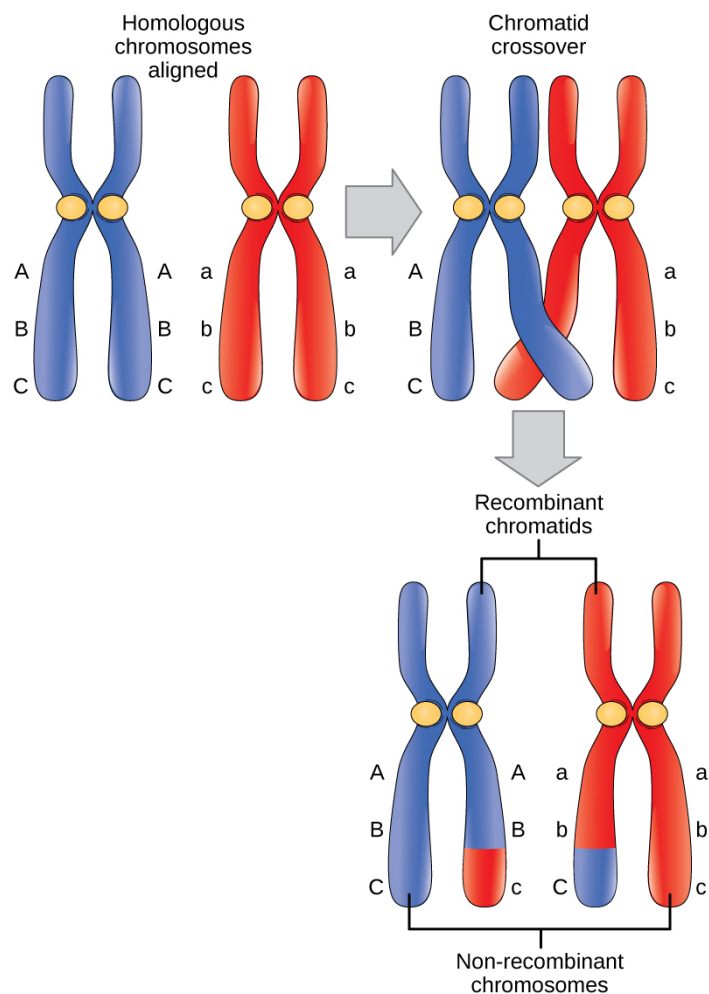


Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

Located at intervals along the synaptonemal complex are large protein assemblies called **recombination nodules**. These assemblies mark the points of later chiasmata and mediate the multistep process of **crossover**—or genetic recombination—between the non-sister chromatids. Near the recombination nodule on each chromatid, the double-stranded DNA is cleaved, the cut ends are modified, and a new connection is made between the non-sister chromatids. As prophase I progresses, the synaptonemal complex begins to break down and the chromosomes begin to condense. When the synaptonemal complex is gone, the homologous chromosomes remain attached to each other at the centromere and at chiasmata. The chiasmata remain until anaphase I. The number of chiasmata varies according to the species and the length of the chromosome. There must be at least one chiasma per chromosome for proper separation of homologous chromosomes during meiosis I, but there may be as many as 25. Following crossover, the synaptonemal complex breaks down and the cohesin connection between homologous pairs is also removed. At the end of prophase I, the pairs are held together only at the chiasmata ([\[link\]](#)) and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation in the nuclei produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete cell it will carry some DNA from one parent of the individual and some DNA from the other parent. The

sister recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover. Multiple crossovers in an arm of the chromosome have the same effect, exchanging segments of DNA to create recombinant chromosomes.



Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes.

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. Kinetochore proteins are multiprotein complexes that bind the centromeres of a chromosome to the microtubules of the mitotic spindle. Microtubules grow from centrosomes placed at opposite poles of the cell. The microtubules move toward the middle of the cell and attach to one of the two fused homologous chromosomes. The microtubules attach at each chromosome's kinetochores. With each member of the homologous pair attached to opposite poles of the cell, in the next phase, the microtubules can pull the homologous pair apart. A spindle fiber that has attached to a kinetochore is called a kinetochore microtubule. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome facing each pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

Metaphase I

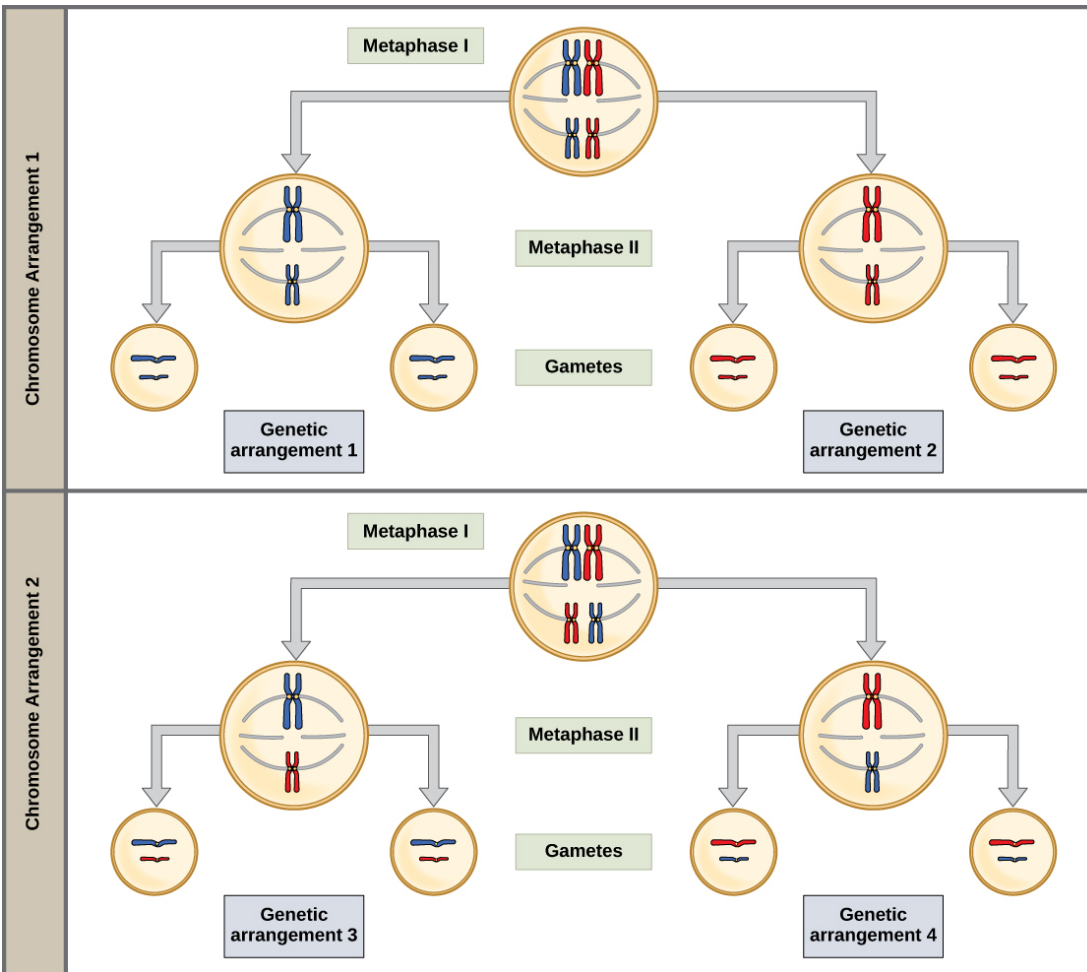
During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The homologous pairs orient themselves randomly at the equator. For example, if the two homologous members of chromosome 1 are labeled a and b, then the chromosomes could line up a-b, or b-a. This is important in determining the genes carried by a gamete, as each will only receive one of the two homologous chromosomes. Recall that homologous chromosomes are not identical. They contain slight differences in their genetic information, causing each gamete to have a unique genetic makeup.

This randomness is the physical basis for the creation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the

egg. Every cell of the multicellular offspring has copies of the original two sets of homologous chromosomes. In prophase I of meiosis, the homologous chromosomes form the tetrads. In metaphase I, these pairs line up at the midway point between the two poles of the cell to form the metaphase plate. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

This event—the random (or independent) assortment of homologous chromosomes at the metaphase plate—is the second mechanism that introduces variation into the gametes or spores. In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations is dependent on the number of chromosomes making up a set. There are two possibilities for orientation at the metaphase plate; the possible number of alignments therefore equals $2n$, where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possible genetically-distinct gametes. This number does not include the variability that was previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition ([link](#)).

To summarize the genetic consequences of meiosis I, the maternal and paternal genes are recombined by crossover events that occur between each homologous pair during prophase I. In addition, the random assortment of tetrads on the metaphase plate produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.



Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes ($n = 2$). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is $2n$, where n equals the number of chromosomes in a set. In this example, there are four possible genetic combinations for the gametes. With $n = 23$ in human cells, there are over 8 million possible combinations of paternal and maternal chromosomes.

Anaphase I

In anaphase I, the microtubules pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. The chiasmata are broken in anaphase I as the microtubules attached to the fused kinetochores pull the homologous chromosomes apart ([link](#)).

Telophase I and Cytokinesis

In telophase, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur, depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I. In other organisms, cytokinesis—the physical separation of the cytoplasmic components into two daughter cells—occurs without reformation of the nuclei. In nearly all species of animals and some fungi, cytokinesis separates the cell contents via a cleavage furrow (constriction of the actin ring that leads to cytoplasmic division). In plants, a cell plate is formed during cell cytokinesis by Golgi vesicles fusing at the metaphase plate. This cell plate will ultimately lead to the formation of cell walls that separate the two daughter cells.

Two haploid cells are the end result of the first meiotic division. The cells are haploid because at each pole, there is just one of each pair of the homologous chromosomes. Therefore, only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though each homolog still consists of two sister chromatids. Recall that sister chromatids are merely duplicates of one of the two homologous chromosomes (except for changes that occurred during crossing over). In meiosis II, these two sister chromatids will separate, creating four haploid daughter cells.

Note:

[Link to Learning](#)



Review the process of meiosis, observing how chromosomes align and migrate, at [Meiosis: An Interactive Animation](#).

Meiosis II

In some species, cells enter a brief interphase, or **interkinesis**, before entering meiosis II. Interkinesis lacks an S phase, so chromosomes are not duplicated. The two cells produced in meiosis I go through the events of meiosis II in synchrony. During meiosis II, the sister chromatids within the two daughter cells separate, forming four new haploid gametes. The mechanics of meiosis II is similar to mitosis, except that each dividing cell has only one set of homologous chromosomes. Therefore, each cell has half the number of sister chromatids to separate out as a diploid cell undergoing mitosis.

Prophase II

If the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes that were duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed.

Prometaphase II

The nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches

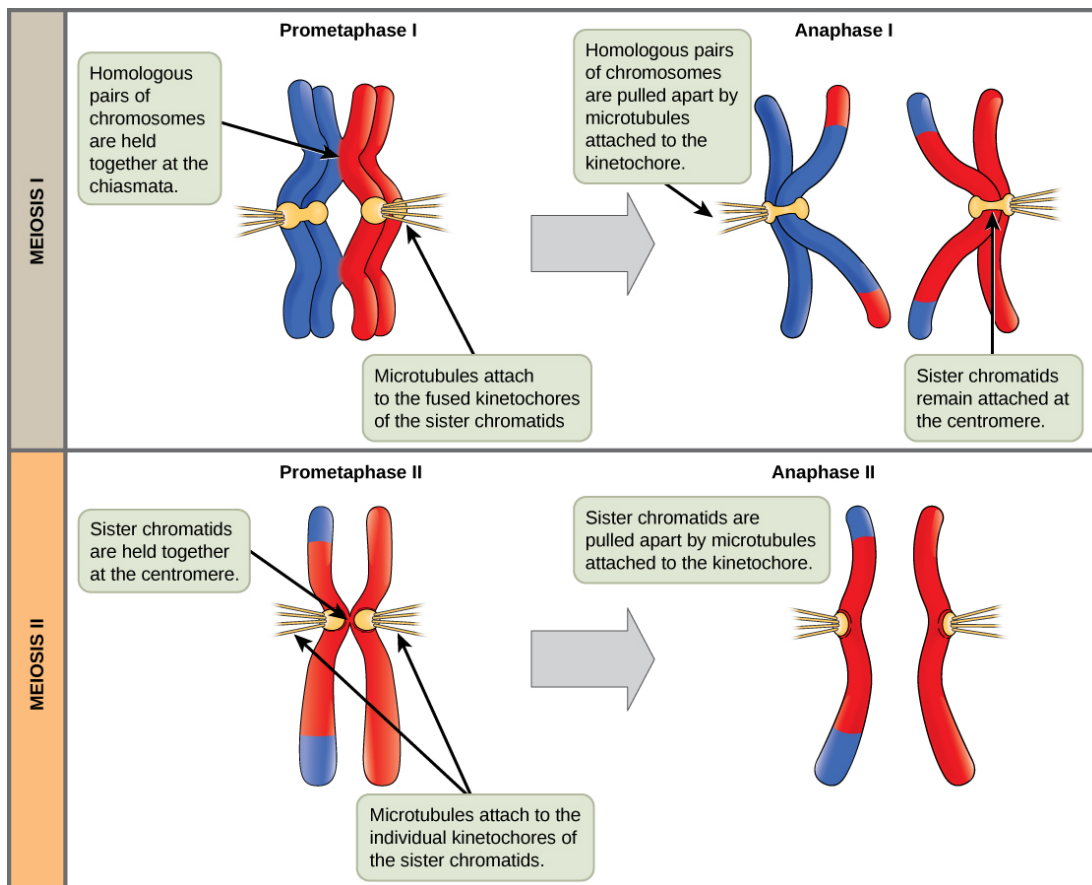
to microtubules from opposite poles.

Metaphase II

The sister chromatids are maximally condensed and aligned at the equator of the cell.

Anaphase II

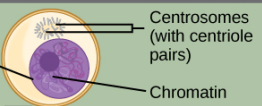
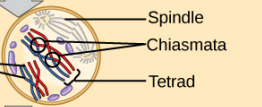
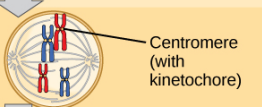
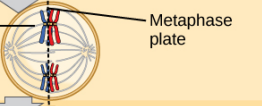
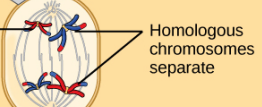


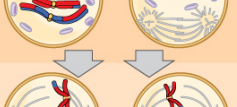
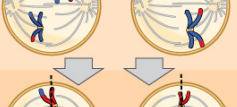
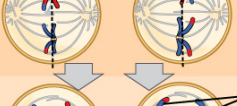

The sister chromatids are pulled apart by the kinetochore microtubules and move toward opposite poles. Non-kinetochore microtubules elongate the cell.



The process of chromosome alignment differs between meiosis I and meiosis II. In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes, and the homologous chromosomes are arranged at the midpoint of the cell in metaphase I. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to the kinetochores of sister chromatids, and the sister chromatids are arranged at the midpoint of the cells in metaphase II. In anaphase II, the sister chromatids are separated.

Telophase II and Cytokinesis

The chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four unique haploid cells. At this point, the newly formed nuclei are both haploid. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombining of maternal and paternal segments of chromosomes (with their sets of genes) that occurs during crossover. The entire process of meiosis is outlined in [\[link\]](#).

Stage	Event	Outcome
INTERPHASE	S phase Nuclear envelope 	Chromosomes are duplicated during interphase. The resulting sister chromatids are held together at the centromere. The centrosomes are also duplicated.
	Prophase I Sister chromatids 	Chromosomes condense, and the nuclear envelope fragments. Homologous chromosomes bind firmly together along their length, forming a tetrad. Chiasmata form between non-sister chromatids. Crossing over occurs at the chiasmata. Spindle fibers emerge from the centrosomes.
MEIOSIS I	Prometaphase I 	Homologous chromosomes are attached to spindle microtubules at the fused kinetochore shared by the sister chromatids. Chromosomes continue to condense, and the nuclear envelope completely disappears.
	Metaphase I Microtubule attached to kinetochore 	Homologous chromosomes randomly assemble at the metaphase plate, where they have been maneuvered into place by the microtubules.
	Anaphase I Sister chromatids remain attached 	Spindle microtubules pull the homologous chromosomes apart. The sister chromatids are still attached at the centromere.
	Telophase I and Cytokinesis 	Sister chromatids arrive at the poles of the cell and begin to decondense. A nuclear envelope forms around each nucleus and the cytoplasm is divided by a cleavage furrow. The result is two haploid cells. Each cell contains one duplicated copy of each homologous chromosome pair.
	Prophase II 	Sister chromatids condense. A new spindle begins to form. The nuclear envelope starts to fragment.
MEIOSIS II	Prometaphase II 	The nuclear envelope disappears, and the spindle fibers engage the individual kinetochores on the sister chromatids.
	Metaphase II 	Sister chromatids line up at the metaphase plate.
	Anaphase II 	Sister chromatids are pulled apart by the shortening of the kinetochore microtubules. Non-kinetochore microtubules lengthen the cell.
	Telophase II and Cytokinesis 	Chromosomes arrive at the poles of the cell and decondense. Nuclear envelopes surround the four nuclei. Cleavage furrows divide the two cells into four haploid cells.
	Haploid daughter cells	

An animal cell with a diploid number of four ($2n = 4$) proceeds through the stages of meiosis to form four haploid daughter cells.

Comparing Meiosis and Mitosis

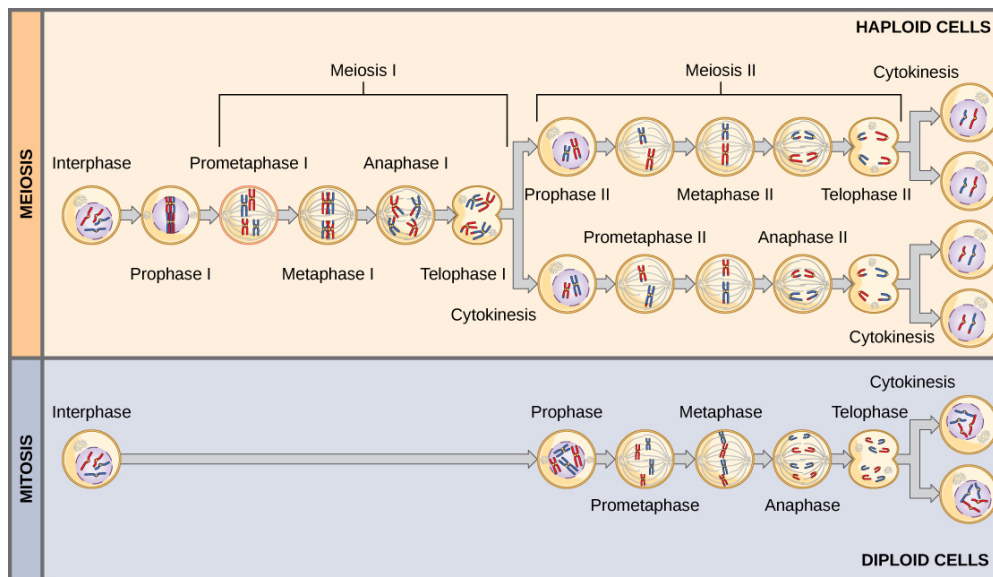
Mitosis and meiosis are both forms of division of the nucleus in eukaryotic cells. They share some similarities, but also exhibit distinct differences that lead to very different outcomes ([link](#)). Mitosis is a single nuclear division that results in two nuclei that are usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original nucleus. They have the same number of sets of chromosomes, one set in the case of haploid cells and two sets in the case of diploid cells. In most plants and all animal species, it is typically diploid cells that undergo mitosis to form new diploid cells. In contrast, meiosis consists of two nuclear divisions resulting in four nuclei that are usually partitioned into four new cells. The nuclei resulting from meiosis are not genetically identical and they contain one chromosome set only. This is half the number of chromosome sets in the original cell, which is diploid.

The main differences between mitosis and meiosis occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together with the synaptonemal complex, develop chiasmata and undergo crossover between sister chromatids, and line up along the metaphase plate in tetrads with kinetochore fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I.

When the chiasmata resolve and the tetrad is broken up with the homologs moving to one pole or another, the ploidy level—the number of sets of chromosomes in each future nucleus—has been reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level during mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, the duplicated chromosomes (only one set of them) line up on the metaphase plate with divided kinetochores attached to kinetochore fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid—now referred to as a chromosome—is pulled to one pole while the other sister chromatid is pulled to the other pole. If it

were not for the fact that there had been crossover, the two products of each individual meiosis II division would be identical (like in mitosis). Instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

Note:**Evolution Connection****The Mystery of the Evolution of Meiosis**

Some characteristics of organisms are so widespread and fundamental that it is sometimes difficult to remember that they evolved like other simpler traits. Meiosis is such an extraordinarily complex series of cellular events that biologists have had trouble hypothesizing and testing how it may have evolved. Although meiosis is inextricably entwined with sexual reproduction and its advantages and disadvantages, it is important to separate the questions of the evolution of meiosis and the evolution of sex, because early meiosis may have been advantageous for different reasons than it is now. Thinking outside the box and imagining what the early benefits from meiosis might have been is one approach to uncovering how it may have evolved.

Meiosis and mitosis share obvious cellular processes and it makes sense that meiosis evolved from mitosis. The difficulty lies in the clear differences between meiosis I and mitosis. Adam Wilkins and Robin Holliday^[footnote] summarized the unique events that needed to occur for the evolution of meiosis from mitosis. These steps are homologous chromosome pairing, crossover exchanges, sister chromatids remaining attached during anaphase, and suppression of DNA replication in interphase. They argue that the first step is the hardest and most important, and that understanding how it evolved would make the evolutionary process clearer. They suggest genetic experiments that might shed light on the evolution of synapsis.

Adam S. Wilkins and Robin Holliday, “The Evolution of Meiosis from Mitosis,” *Genetics* 181 (2009): 3–12.

There are other approaches to understanding the evolution of meiosis in progress. Different forms of meiosis exist in single-celled protists. Some appear to be simpler or more “primitive” forms of meiosis. Comparing the meiotic divisions of different protists may shed light on the evolution of meiosis. Marilee Ramesh and colleagues^[footnote] compared the genes involved in meiosis in protists to understand when and where meiosis might have evolved. Although research is still ongoing, recent scholarship

into meiosis in protists suggests that some aspects of meiosis may have evolved later than others. This kind of genetic comparison can tell us what aspects of meiosis are the oldest and what cellular processes they may have borrowed from in earlier cells.

Marilee A. Ramesh, Shehre-Banoo Malik and John M. Logsdon, Jr, “A Phylogenetic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis,” *Current Biology* 15 (2005):185–91.

Note:

Link to Learning



Click through the steps of this interactive animation to compare the meiotic process of cell division to that of mitosis: [How Cells Divide](#).

Section Summary

Sexual reproduction requires that diploid organisms produce haploid cells that can fuse during fertilization to form diploid offspring. As with mitosis, DNA replication occurs prior to meiosis during the S-phase of the cell cycle. Meiosis is a series of events that arrange and separate chromosomes and chromatids into daughter cells. During the interphases of meiosis, each chromosome is duplicated. In meiosis, there are two rounds of nuclear division resulting in four nuclei and usually four daughter cells, each with half the number of chromosomes as the parent cell. The first separates homologs, and the second—like mitosis—separates chromatids into individual chromosomes. During meiosis, variation in the daughter nuclei is

introduced because of crossover in prophase I and random alignment of tetrads at metaphase I. The cells that are produced by meiosis are genetically unique.

Meiosis and mitosis share similarities, but have distinct outcomes. Mitotic divisions are single nuclear divisions that produce daughter nuclei that are genetically identical and have the same number of chromosome sets as the original cell. Meiotic divisions include two nuclear divisions that produce four daughter nuclei that are genetically different and have one chromosome set instead of the two sets of chromosomes in the parent cell. The main differences between the processes occur in the first division of meiosis, in which homologous chromosomes are paired and exchange non-sister chromatid segments. The homologous chromosomes separate into different nuclei during meiosis I, causing a reduction of ploidy level in the first division. The second division of meiosis is more similar to a mitotic division, except that the daughter cells do not contain identical genomes because of crossover.

Review Questions

Exercise:

Problem: Meiosis produces _____ daughter cells.

- a. two haploid
- b. two diploid
- c. four haploid
- d. four diploid

Solution:

C

Exercise:

Problem: What structure is most important in forming the tetrads?

- a. centromere
- b. synaptonemal complex
- c. chiasma
- d. kinetochore

Solution:

B

Exercise:

Problem:

At which stage of meiosis are sister chromatids separated from each other?

- a. prophase I
- b. prophase II
- c. anaphase I
- d. anaphase II

Solution:

D

Exercise:

Problem:

At metaphase I, homologous chromosomes are connected only at what structures?

- a. chiasmata
- b. recombination nodules
- c. microtubules
- d. kinetochores

Solution:

A

Exercise:

Problem: Which of the following is *not* true in regard to crossover?

- a. Spindle microtubules guide the transfer of DNA across the synaptonemal complex.
- b. Non-sister chromatids exchange genetic material.
- c. Chiasmata are formed.
- d. Recombination nodules mark the crossover point.

Solution:

C

Exercise:

Problem:

What phase of mitotic interphase is missing from meiotic interkinesis?

- a. G₀ phase
- b. G₁ phase
- c. S phase
- d. G₂ phase

Solution:

C

Exercise:

Problem: The part of meiosis that is similar to mitosis is _____.

- a. meiosis I
- b. anaphase I
- c. meiosis II

d. interkinesis

Solution:

C

Exercise:

Problem:

If a muscle cell of a typical organism has 32 chromosomes, how many chromosomes will be in a gamete of that same organism?

- a. 8
- b. 16
- c. 32
- d. 64

Solution:

B

Free Response

Exercise:

Problem: Describe the process that results in the formation of a tetrad.

Solution:

During the meiotic interphase, each chromosome is duplicated. The sister chromatids that are formed during synthesis are held together at the centromere region by cohesin proteins. All chromosomes are attached to the nuclear envelope by their tips. As the cell enters prophase I, the nuclear envelope begins to fragment, and the proteins holding homologous chromosomes locate each other. The four sister

chromatids align lengthwise, and a protein lattice called the synaptonemal complex is formed between them to bind them together. The synaptonemal complex facilitates crossover between non-sister chromatids, which is observed as chiasmata along the length of the chromosome. As prophase I progresses, the synaptonemal complex breaks down and the sister chromatids become free, except where they are attached by chiasmata. At this stage, the four chromatids are visible in each homologous pairing and are called a tetrad.

Exercise:

Problem:

Explain how the random alignment of homologous chromosomes during metaphase I contributes to the variation in gametes produced by meiosis.

Solution:

Random alignment leads to new combinations of traits. The chromosomes that were originally inherited by the gamete-producing individual came equally from the egg and the sperm. In metaphase I, the duplicated copies of these maternal and paternal homologous chromosomes line up across the center of the cell. The orientation of each tetrad is random. There is an equal chance that the maternally derived chromosomes will be facing either pole. The same is true of the paternally derived chromosomes. The alignment should occur differently in almost every meiosis. As the homologous chromosomes are pulled apart in anaphase I, any combination of maternal and paternal chromosomes will move toward each pole. The gametes formed from these two groups of chromosomes will have a mixture of traits from the individual's parents. Each gamete is unique.

Exercise:

Problem:

What is the function of the fused kinetochore found on sister chromatids in prometaphase I?

Solution:

In metaphase I, the homologous chromosomes line up at the metaphase plate. In anaphase I, the homologous chromosomes are pulled apart and move to opposite poles. Sister chromatids are not separated until meiosis II. The fused kinetochore formed during meiosis I ensures that each spindle microtubule that binds to the tetrad will attach to both sister chromatids.

Exercise:**Problem:**

In a comparison of the stages of meiosis to the stages of mitosis, which stages are unique to meiosis and which stages have the same events in both meiosis and mitosis?

Solution:

All of the stages of meiosis I, except possibly telophase I, are unique because homologous chromosomes are separated, not sister chromatids. In some species, the chromosomes do not decondense and the nuclear envelopes do not form in telophase I. All of the stages of meiosis II have the same events as the stages of mitosis, with the possible exception of prophase II. In some species, the chromosomes are still condensed and there is no nuclear envelope. Other than this, all processes are the same.

Glossary**chiasmata**

(singular, *chiasma*) the structure that forms at the crossover points after genetic material is exchanged

cohesin

proteins that form a complex that seals sister chromatids together at their centromeres until anaphase II of meiosis

crossover

exchange of genetic material between non-sister chromatids resulting in chromosomes that incorporate genes from both parents of the organism

fertilization

union of two haploid cells from two individual organisms

interkinesis

(also, *interphase II*) brief period of rest between meiosis I and meiosis II

meiosis

a nuclear division process that results in four haploid cells

meiosis I

first round of meiotic cell division; referred to as reduction division because the ploidy level is reduced from diploid to haploid

meiosis II

second round of meiotic cell division following meiosis I; sister chromatids are separated into individual chromosomes, and the result is four unique haploid cells

recombination nodules

protein assemblies formed on the synaptonemal complex that mark the points of crossover events and mediate the multistep process of genetic recombination between non-sister chromatids

reduction division

nuclear division that produces daughter nuclei each having one-half as many chromosome sets as the parental nucleus; meiosis I is a reduction division

somatic cell

all the cells of a multicellular organism except the gametes or reproductive cells

spore

haploid cell that can produce a haploid multicellular organism or can fuse with another spore to form a diploid cell

synapsis

formation of a close association between homologous chromosomes during prophase I

synaptonemal complex

protein lattice that forms between homologous chromosomes during prophase I, supporting crossover

tetrad

two duplicated homologous chromosomes (four chromatids) bound together by chiasmata during prophase I

Sexual Reproduction

By the end of this section, you will be able to:

- Explain that meiosis and sexual reproduction are evolved traits
- Identify variation among offspring as a potential evolutionary advantage to sexual reproduction
- Describe the three different life-cycle types among sexual multicellular organisms and their commonalities

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. It appears to have been very successful because most eukaryotes are able to reproduce sexually, and in many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, creating offspring that are genetic clones of the parent appears to be a better system. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring whenever circumstances are favorable by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. Indeed, some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, in asexual populations, every individual is capable of reproduction. In sexual populations, the males are not producing the offspring themselves, so in theory an asexual population could grow twice as fast.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexuality (and meiosis) so common? This is one of the important unanswered questions in biology and has been the focus of much research beginning in the latter half of the twentieth century. There are several possible explanations, one of which is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of the population. Thus, on average, a sexually reproducing population will leave more descendants than an otherwise similar asexually reproducing population. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms, but in addition, those different

mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes and the genes are mixed into different combinations by crossovers during prophase I and random assortment at metaphase I.

Note:

Evolution Connection

The Red Queen Hypothesis

It is not in dispute that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. But why, even in the face of fairly stable conditions, does sexual reproduction persist when it is more difficult and costly for individual organisms? Variation is the outcome of sexual reproduction, but why are ongoing variations necessary? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973. [\[footnote\]](#) The concept was named in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass*.

Leigh Van Valen, "A New Evolutionary Law," *Evolutionary Theory* 1 (1973): 1–30

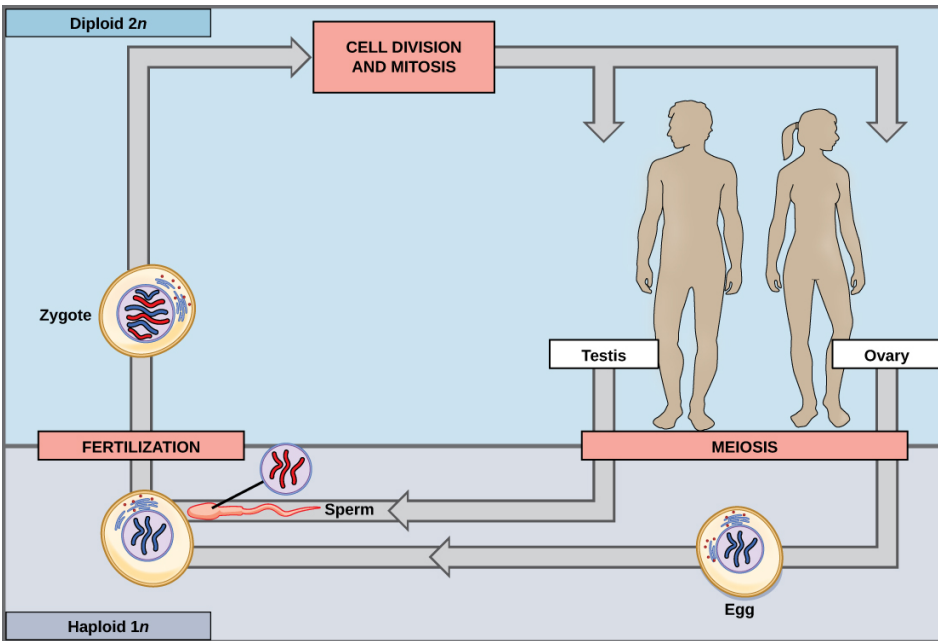
All species co-evolve with other organisms; for example predators evolve with their prey, and parasites evolve with their hosts. Each tiny advantage gained by favorable variation gives a species an edge over close competitors, predators, parasites, or even prey. The only method that will allow a co-evolving species to maintain its own share of the resources is to also continually improve its fitness. As one species gains an advantage, this increases selection on the other species; they must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among the progeny of sexual reproduction provides all species with a mechanism to improve rapidly. Species that cannot keep up become extinct. The Red Queen's catchphrase was, "It takes all the running you can do to stay in the same place." This is an apt description of co-evolution between competing species.

Life Cycles of Sexually Reproducing Organisms

Fertilization and meiosis alternate in sexual **life cycles**. What happens between these two events depends on the organism. The process of meiosis reduces the chromosome number by half. Fertilization, the joining of two haploid gametes, restores the diploid condition. There are three main categories of life cycles in multicellular organisms: **diploid-dominant**, in which the multicellular diploid stage is the most obvious life stage, such as with most animals including humans; **haploid-dominant**, in which the multicellular haploid stage is the most obvious life stage, such as with all fungi and some algae; and **alternation of generations**, in which the two stages are apparent to different degrees depending on the group, as with plants and some algae.

Diploid-Dominant Life Cycle

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism are the gametes. Early in the development of the embryo, specialized diploid cells, called **germ cells**, are produced within the gonads, such as the testes and ovaries. Germ cells are capable of mitosis to perpetuate the cell line and meiosis to produce gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state ([link](#)).



In animals, sexually reproducing adults form haploid gametes from diploid germ cells. Fusion of the gametes gives rise to a fertilized egg cell, or zygote. The zygote will undergo multiple rounds of mitosis to produce a multicellular offspring. The germ cells are generated early in the development of the zygote.

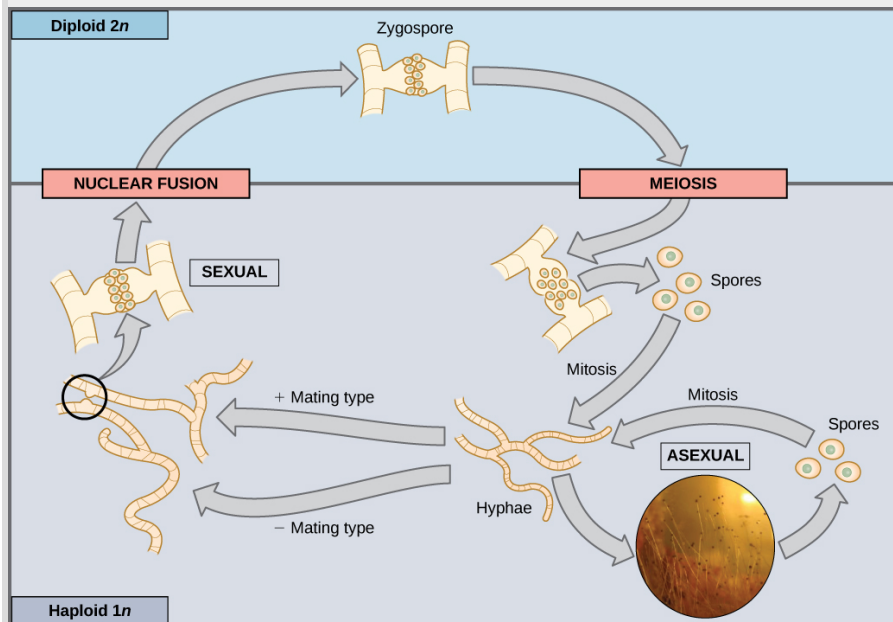
Haploid-Dominant Life Cycle

Most fungi and algae employ a life-cycle type in which the “body” of the organism—the ecologically important part of the life cycle—is haploid. The haploid cells that make up the tissues of the dominant multicellular stage are formed by mitosis. During sexual reproduction, specialized haploid cells from two individuals, designated the (+) and (−) mating types, join to form a diploid zygote. The zygote immediately undergoes meiosis to form four haploid cells called spores. Although haploid like the “parents,” these spores contain a new genetic combination from two parents. The spores can remain dormant for various time periods. Eventually, when conditions are

conductive, the spores form multicellular haploid structures by many rounds of mitosis ([\[link\]](#)).

Note:

Art Connection

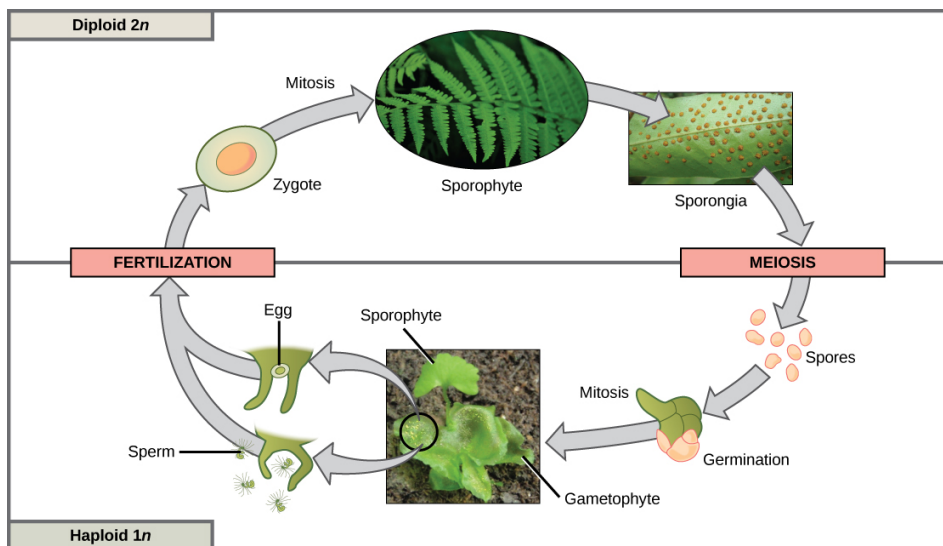


Fungi, such as black bread mold (*Rhizopus nigricans*), have haploid-dominant life cycles. The haploid multicellular stage produces specialized haploid cells by mitosis that fuse to form a diploid zygote. The zygote undergoes meiosis to produce haploid spores. Each spore gives rise to a multicellular haploid organism by mitosis. (credit “zygomycota” micrograph: modification of work by “Fanaberka”/Wikimedia Commons)

If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

Alternation of Generations

The third life-cycle type, employed by some algae and all plants, is a blend of the haploid-dominant and diploid-dominant extremes. Species with alternation of generations have both haploid and diploid multicellular organisms as part of their life cycle. The haploid multicellular plants are called **gametophytes**, because they produce gametes from specialized cells. Meiosis is not directly involved in the production of gametes in this case, because the organism that produces the gametes is already a haploid. Fertilization between the gametes forms a diploid zygote. The zygote will undergo many rounds of mitosis and give rise to a diploid multicellular plant called a **sporophyte**. Specialized cells of the sporophyte will undergo meiosis and produce haploid spores. The spores will subsequently develop into the gametophytes ([\[link\]](#)).



Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. In some plants, such as ferns, both the haploid and diploid plant stages are free-living. The diploid plant is called a sporophyte because it produces haploid spores by meiosis. The spores develop into multicellular, haploid plants called gametophytes because they produce gametes. The

gametes of two individuals will fuse to form a diploid zygote that becomes the sporophyte. (credit “fern”: modification of work by Cory Zanker; credit “sporangia”: modification of work by "Obsidian Soul"/Wikimedia Commons; credit “gametophyte and sporophyte”: modification of work by “Vlmastra”/Wikimedia Commons)

Although all plants utilize some version of the alternation of generations, the relative size of the sporophyte and the gametophyte and the relationship between them vary greatly. In plants such as moss, the gametophyte organism is the free-living plant, and the sporophyte is physically dependent on the gametophyte. In other plants, such as ferns, both the gametophyte and sporophyte plants are free-living; however, the sporophyte is much larger. In seed plants, such as magnolia trees and daisies, the gametophyte is composed of only a few cells and, in the case of the female gametophyte, is completely retained within the sporophyte.

Sexual reproduction takes many forms in multicellular organisms. However, at some point in each type of life cycle, meiosis produces haploid cells that will fuse with the haploid cell of another organism. The mechanisms of variation—crossover, random assortment of homologous chromosomes, and random fertilization—are present in all versions of sexual reproduction. The fact that nearly every multicellular organism on Earth employs sexual reproduction is strong evidence for the benefits of producing offspring with unique gene combinations, though there are other possible benefits as well.

Section Summary

Nearly all eukaryotes undergo sexual reproduction. The variation introduced into the reproductive cells by meiosis appears to be one of the advantages of sexual reproduction that has made it so successful. Meiosis and fertilization alternate in sexual life cycles. The process of meiosis produces unique reproductive cells called gametes, which have half the

number of chromosomes as the parent cell. Fertilization, the fusion of haploid gametes from two individuals, restores the diploid condition. Thus, sexually reproducing organisms alternate between haploid and diploid stages. However, the ways in which reproductive cells are produced and the timing between meiosis and fertilization vary greatly. There are three main categories of life cycles: diploid-dominant, demonstrated by most animals; haploid-dominant, demonstrated by all fungi and some algae; and the alternation of generations, demonstrated by plants and some algae.

Art Connections

Exercise:

Problem:

[\[link\]](#) If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

Solution:

[\[link\]](#) Yes, it will be able to reproduce asexually.

Review Questions

Exercise:

Problem:

What is a likely evolutionary advantage of sexual reproduction over asexual reproduction?

- a. Sexual reproduction involves fewer steps.
- b. There is a lower chance of using up the resources in a given environment.
- c. Sexual reproduction results in variation in the offspring.
- d. Sexual reproduction is more cost-effective.

Solution:

C

Exercise:

Problem:

Which type of life cycle has both a haploid and diploid multicellular stage?

- a. asexual
- b. diploid-dominant
- c. haploid-dominant
- d. alternation of generations

Solution:

D

Exercise:

Problem: Fungi typically display which type of life cycle?

- a. diploid-dominant
- b. haploid-dominant
- c. alternation of generations
- d. asexual

Solution:

B

Exercise:

Problem:

A diploid, multicellular life-cycle stage that gives rise to haploid cells by meiosis is called a _____.

- a. sporophyte
- b. gametophyte
- c. spore
- d. gamete

Solution:

A

Free Response**Exercise:****Problem:**

List and briefly describe the three processes that lead to variation in offspring with the same parents.

Solution:

a. Crossover occurs in prophase I between non-sister homologous chromosomes. Segments of DNA are exchanged between maternally derived and paternally derived chromosomes, and new gene combinations are formed. b. Random alignment during metaphase I leads to gametes that have a mixture of maternal and paternal chromosomes. c. Fertilization is random, in that any two gametes can fuse.

Exercise:

Problem:

Compare the three main types of life cycles in multicellular organisms and give an example of an organism that employs each.

Solution:

a. In the haploid-dominant life cycle, the multicellular stage is haploid. The diploid stage is a spore that undergoes meiosis to produce cells that will divide mitotically to produce new multicellular organisms. Fungi have a haploid-dominant life cycle. b. In the diploid-dominant life cycle, the most visible or largest multicellular stage is diploid. The haploid stage is usually reduced to a single cell type, such as a gamete or spore. Animals, such as humans, have a diploid-dominant life cycle. c. In the alternation of generations life cycle, there are both haploid and diploid multicellular stages, although the haploid stage may be completely retained by the diploid stage. Plants have a life cycle with alternation of generations.

Glossary

alternation of generations

life-cycle type in which the diploid and haploid stages alternate

diploid-dominant

life-cycle type in which the multicellular diploid stage is prevalent

haploid-dominant

life-cycle type in which the multicellular haploid stage is prevalent

gametophyte

a multicellular haploid life-cycle stage that produces gametes

germ cells

specialized cell line that produces gametes, such as eggs or sperm

life cycle

the sequence of events in the development of an organism and the production of cells that produce offspring

sporophyte

a multicellular diploid life-cycle stage that produces haploid spores by meiosis

Introduction

class="introduction"

Experimentin
g with
thousands of
garden peas,
Mendel
uncovered the
fundamentals
of genetics.
(credit:
modification
of work by
Jerry
Kirkhart)



Genetics is the study of heredity. Johann Gregor Mendel set the framework for genetics long before chromosomes or genes had been identified, at a time when meiosis was not well understood. Mendel selected a simple biological system and conducted methodical, quantitative analyses using

large sample sizes. Because of Mendel's work, the fundamental principles of heredity were revealed. We now know that genes, carried on chromosomes, are the basic functional units of heredity with the capability to be replicated, expressed, or mutated. Today, the postulates put forth by Mendel form the basis of classical, or Mendelian, genetics. Not all genes are transmitted from parents to offspring according to Mendelian genetics, but Mendel's experiments serve as an excellent starting point for thinking about inheritance.

Mendel's Experiments and the Laws of Probability

By the end of this section, you will be able to:

- Describe the scientific reasons for the success of Mendel's experimental work
- Describe the expected outcomes of monohybrid crosses involving dominant and recessive alleles
- Apply the sum and product rules to calculate probabilities



Johann Gregor Mendel is considered the father of genetics.

Johann Gregor Mendel (1822–1884) ([\[link\]](#)) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance patterns in honeybees and plants, ultimately settling on pea plants as his primary **model system** (a system with convenient characteristics used to study a specific biological phenomenon to be applied to other systems). In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local Natural History Society. He demonstrated that traits are transmitted faithfully from parents to offspring independently of other traits and in dominant and recessive patterns. In 1866, he published his work, *Experiments in Plant Hybridization*, [\[footnote\]](#) in the proceedings of the Natural History Society of Brünn.

Johann Gregor Mendel, *Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865 Abhandlungen*, 3–47.

[for English translation see <http://www.mendelweb.org/Mendel.plain.html>]

Mendel's work went virtually unnoticed by the scientific community that believed, incorrectly, that the process of inheritance involved a blending of parental traits that produced an intermediate physical appearance in offspring; this hypothetical process appeared to be correct because of what we know now as continuous variation.

Continuous variation results from the action of many genes to determine a characteristic like human height. Offspring appear to be a “blend” of their parents' traits when we look at characteristics that exhibit continuous variation. The **blending theory of inheritance** asserted that the original parental traits were lost or absorbed by the blending in the offspring, but we now know that this is not the case. Mendel was the first researcher to see it. Instead of continuous characteristics, Mendel worked with traits that were inherited in distinct classes (specifically, violet versus white flowers); this is referred to as **discontinuous variation**. Mendel's choice of these kinds of traits allowed him to see experimentally that the traits were not blended in the offspring, nor were they absorbed, but rather that they kept their distinctness and could be passed on. In 1868, Mendel became abbot of the monastery and exchanged his scientific pursuits for his pastoral duties. He was not recognized for his extraordinary scientific contributions during his lifetime. In fact, it was not until 1900 that his work was rediscovered, reproduced, and revitalized by scientists on the brink of discovering the chromosomal basis of heredity.

Mendel's Model System

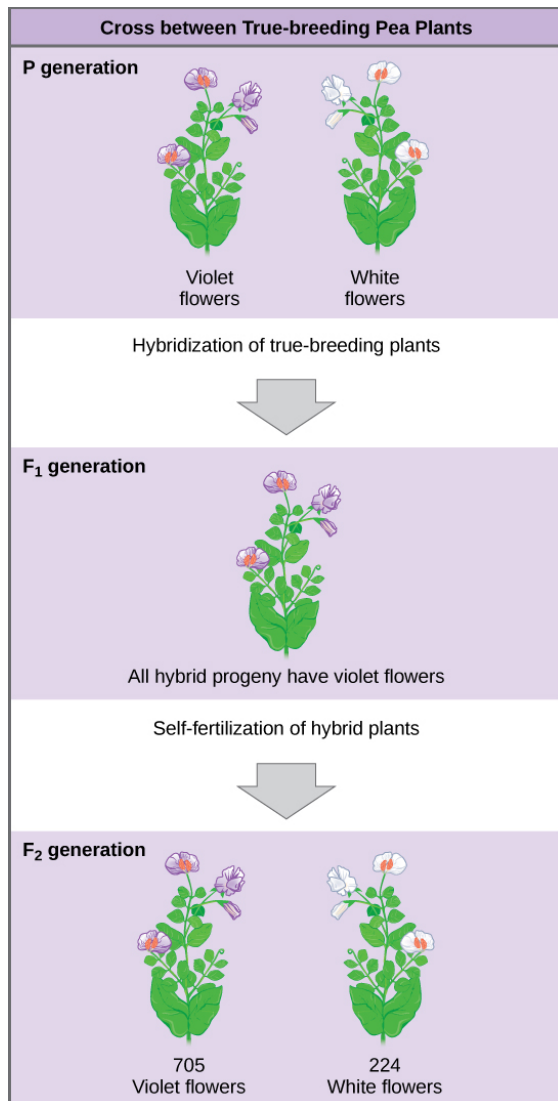
Mendel's seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, such that pollen encounters ova within individual flowers. The flower petals remain sealed tightly until after pollination, preventing pollination from other plants. The result is highly inbred, or “true-breeding,” pea plants. These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding. The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

Mendelian Crosses

Mendel performed **hybridizations**, which involve mating two true-breeding individuals that have different traits. In the pea, which is naturally self-pollinating, this is done by manually transferring pollen from the anther of a mature pea plant of one variety to the stigma of a separate mature pea plant of the second variety. In plants, pollen carries the

male gametes (sperm) to the stigma, a sticky organ that traps pollen and allows the sperm to move down the pistil to the female gametes (ova) below. To prevent the pea plant that was receiving pollen from self-fertilizing and confounding his results, Mendel painstakingly removed all of the anthers from the plant's flowers before they had a chance to mature.

Plants used in first-generation crosses were called **P₀**, or parental generation one, plants ([link](#)). Mendel collected the seeds belonging to the P₀ plants that resulted from each cross and grew them the following season. These offspring were called the **F₁**, or the first filial (*filial* = offspring, daughter or son), generation. Once Mendel examined the characteristics in the F₁ generation of plants, he allowed them to self-fertilize naturally. He then collected and grew the seeds from the F₁ plants to produce the **F₂**, or second filial, generation. Mendel's experiments extended beyond the F₂ generation to the F₃ and F₄ generations, and so on, but it was the ratio of characteristics in the P₀-F₁-F₂ generations that were the most intriguing and became the basis for Mendel's postulates.



In one of his experiments on inheritance patterns, Mendel crossed plants that were true-breeding for violet flower color with plants true-breeding for white flower color (the P generation).

The resulting hybrids in the F₁ generation all had violet flowers.

In the F₂ generation, approximately three quarters of the plants had violet flowers, and one quarter had white flowers.

Garden Pea Characteristics Revealed the Basics of Heredity

In his 1865 publication, Mendel reported the results of his crosses involving seven different characteristics, each with two contrasting traits. A **trait** is defined as a variation in the physical appearance of a heritable characteristic. The characteristics included plant height, seed texture, seed color, flower color, pea pod size, pea pod color, and flower position. For the characteristic of flower color, for example, the two contrasting traits were white versus violet. To fully examine each characteristic, Mendel generated large numbers of F_1 and F_2 plants, reporting results from 19,959 F_2 plants alone. His findings were consistent.

What results did Mendel find in his crosses for flower color? First, Mendel confirmed that he had plants that bred true for white or violet flower color. Regardless of how many generations Mendel examined, all self-crossed offspring of parents with white flowers had white flowers, and all self-crossed offspring of parents with violet flowers had violet flowers. In addition, Mendel confirmed that, other than flower color, the pea plants were physically identical.

Once these validations were complete, Mendel applied the pollen from a plant with violet flowers to the stigma of a plant with white flowers. After gathering and sowing the seeds that resulted from this cross, Mendel found that 100 percent of the F_1 hybrid generation had violet flowers. Conventional wisdom at that time would have predicted the hybrid flowers to be pale violet or for hybrid plants to have equal numbers of white and violet flowers. In other words, the contrasting parental traits were expected to blend in the offspring. Instead, Mendel's results demonstrated that the white flower trait in the F_1 generation had completely disappeared.

Importantly, Mendel did not stop his experimentation there. He allowed the F_1 plants to self-fertilize and found that, of F_2 -generation plants, 705 had violet flowers and 224 had white flowers. This was a ratio of 3.15 violet flowers per one white flower, or approximately 3:1. When Mendel transferred pollen from a plant with violet flowers to the stigma of a plant with white flowers and vice versa, he obtained about the same ratio regardless of which parent, male or female, contributed which trait. This is called a **reciprocal cross**—a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross. For the other six characteristics Mendel examined, the F_1 and F_2 generations behaved in the same way as they had for flower color. One of the two traits would disappear completely from the F_1 generation only to reappear in the F_2 generation at a ratio of approximately 3:1 ([link](#)).

The Results of Mendel's Garden Pea Hybridizations				
Characteristic	Contrasting P ₀ Traits	F ₁ Offspring Traits	F ₂ Offspring Traits	F ₂ Trait Ratios
Flower color	Violet vs. white	100 percent violet	705 violet 224 white	3.15:1
Flower position	Axial vs. terminal	100 percent axial	651 axial 207 terminal	3.14:1
Plant height	Tall vs. dwarf	100 percent tall	787 tall 277 dwarf	2.84:1
Seed texture	Round vs. wrinkled	100 percent round	5,474 round 1,850 wrinkled	2.96:1
Seed color	Yellow vs. green	100 percent yellow	6,022 yellow 2,001 green	3.01:1
Pea pod texture	Inflated vs. constricted	100 percent inflated	882 inflated 299 constricted	2.95:1

The Results of Mendel's Garden Pea Hybridizations				
Characteristic	Contrasting P ₀ Traits	F ₁ Offspring Traits	F ₂ Offspring Traits	F ₂ Trait Ratios
Pea pod color	Green vs. yellow	100 percent green	428 green 152 yellow	2.82:1

Upon compiling his results for many thousands of plants, Mendel concluded that the characteristics could be divided into expressed and latent traits. He called these, respectively, dominant and recessive traits. **Dominant traits** are those that are inherited unchanged in a hybridization. **Recessive traits** become latent, or disappear, in the offspring of a hybridization. The recessive trait does, however, reappear in the progeny of the hybrid offspring. An example of a dominant trait is the violet-flower trait. For this same characteristic (flower color), white-colored flowers are a recessive trait. The fact that the recessive trait reappeared in the F₂ generation meant that the traits remained separate (not blended) in the plants of the F₁ generation. Mendel also proposed that plants possessed two copies of the trait for the flower-color characteristic, and that each parent transmitted one of its two copies to its offspring, where they came together. Moreover, the physical observation of a dominant trait could mean that the genetic composition of the organism included two dominant versions of the characteristic or that it included one dominant and one recessive version. Conversely, the observation of a recessive trait meant that the organism lacked any dominant versions of this characteristic.

So why did Mendel repeatedly obtain 3:1 ratios in his crosses? To understand how Mendel deduced the basic mechanisms of inheritance that lead to such ratios, we must first review the laws of probability.

Probability Basics

Probabilities are mathematical measures of likelihood. The empirical probability of an event is calculated by dividing the number of times the event occurs by the total number of opportunities for the event to occur. It is also possible to calculate theoretical probabilities by dividing the number of times that an event is expected to occur by the number of times that it could occur. Empirical probabilities come from observations, like those of Mendel. Theoretical probabilities come from knowing how the events are produced and assuming that the probabilities of individual outcomes are equal. A

probability of one for some event indicates that it is guaranteed to occur, whereas a probability of zero indicates that it is guaranteed not to occur. An example of a genetic event is a round seed produced by a pea plant. In his experiment, Mendel demonstrated that the probability of the event “round seed” occurring was one in the F_1 offspring of true-breeding parents, one of which has round seeds and one of which has wrinkled seeds. When the F_1 plants were subsequently self-crossed, the probability of any given F_2 offspring having round seeds was now three out of four. In other words, in a large population of F_2 offspring chosen at random, 75 percent were expected to have round seeds, whereas 25 percent were expected to have wrinkled seeds. Using large numbers of crosses, Mendel was able to calculate probabilities and use these to predict the outcomes of other crosses.

The Product Rule and Sum Rule

Mendel demonstrated that the pea-plant characteristics he studied were transmitted as discrete units from parent to offspring. As will be discussed, Mendel also determined that different characteristics, like seed color and seed texture, were transmitted independently of one another and could be considered in separate probability analyses. For instance, performing a cross between a plant with green, wrinkled seeds and a plant with yellow, round seeds still produced offspring that had a 3:1 ratio of green:yellow seeds (ignoring seed texture) and a 3:1 ratio of round:wrinkled seeds (ignoring seed color). The characteristics of color and texture did not influence each other.

The **product rule** of probability can be applied to this phenomenon of the independent transmission of characteristics. The product rule states that the probability of two independent events occurring together can be calculated by multiplying the individual probabilities of each event occurring alone. To demonstrate the product rule, imagine that you are rolling a six-sided die (D) and flipping a penny (P) at the same time. The die may roll any number from 1–6 ($D_{\#}$), whereas the penny may turn up heads (P_H) or tails (P_T). The outcome of rolling the die has no effect on the outcome of flipping the penny and vice versa. There are 12 possible outcomes of this action ([\[link\]](#)), and each event is expected to occur with equal probability.

Twelve Equally Likely Outcomes of Rolling a Die and Flipping a Penny	
Rolling Die	Flipping Penny

Twelve Equally Likely Outcomes of Rolling a Die and Flipping a Penny	
Rolling Die	Flipping Penny
D ₁	P _H
D ₁	P _T
D ₂	P _H
D ₂	P _T
D ₃	P _H
D ₃	P _T
D ₄	P _H
D ₄	P _T
D ₅	P _H
D ₅	P _T
D ₆	P _H
D ₆	P _T

Of the 12 possible outcomes, the die has a 2/12 (or 1/6) probability of rolling a two, and the penny has a 6/12 (or 1/2) probability of coming up heads. By the product rule, the probability that you will obtain the combined outcome 2 and heads is: (D₂) x (P_H) = (1/6) x (1/2) or 1/12 ([link](#)). Notice the word “and” in the description of the probability. The “and” is a signal to apply the product rule. For example, consider how the product rule is applied to the dihybrid cross: the probability of having both dominant traits in the F₂ progeny is the product of the probabilities of having the dominant trait for each characteristic, as shown here:

Equation:

$$\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$$

On the other hand, the **sum rule** of probability is applied when considering two mutually exclusive outcomes that can come about by more than one pathway. The sum rule states that the probability of the occurrence of one event or the other event, of two mutually exclusive events, is the sum of their individual probabilities. Notice the word “or” in the description of the probability. The “or” indicates that you should apply the sum rule. In this case, let’s imagine you are flipping a penny (P) and a quarter (Q). What is the probability of one coin coming up heads and one coin coming up tails? This outcome can be achieved by two cases: the penny may be heads (P_H) and the quarter may be tails (Q_T), or the quarter may be heads (Q_H) and the penny may be tails (P_T). Either case fulfills the outcome. By the sum rule, we calculate the probability of obtaining one head and one tail as [(P_H) × (Q_T)] + [(Q_H) × (P_T)] = [(1/2) × (1/2)] + [(1/2) × (1/2)] = 1/2 ([link](#)). You should also notice that we used the product rule to calculate the probability of P_H and Q_T, and also the probability of P_T and Q_H, before we summed them. Again, the sum rule can be applied to show the probability of having just one dominant trait in the F₂ generation of a dihybrid cross:

Equation:

$$\frac{3}{16} + \frac{3}{4} = \frac{15}{16}$$

The Product Rule and Sum Rule	
Product Rule	Sum Rule
For independent events A and B, the probability (P) of them both occurring (<i>A and B</i>) is (P _A × P _B)	For mutually exclusive events A and B, the probability (P) that at least one occurs (<i>A or B</i>) is (P _A + P _B)

To use probability laws in practice, it is necessary to work with large sample sizes because small sample sizes are prone to deviations caused by chance. The large quantities of pea plants that Mendel examined allowed him calculate the probabilities of the traits appearing in his F₂ generation. As you will learn, this discovery meant that when parental traits were known, the offspring’s traits could be predicted accurately even before fertilization.

Section Summary

Working with garden pea plants, Mendel found that crosses between parents that differed by one trait produced F_1 offspring that all expressed the traits of one parent. Observable traits are referred to as dominant, and non-expressed traits are described as recessive. When the offspring in Mendel's experiment were self-crossed, the F_2 offspring exhibited the dominant trait or the recessive trait in a 3:1 ratio, confirming that the recessive trait had been transmitted faithfully from the original P_0 parent. Reciprocal crosses generated identical F_1 and F_2 offspring ratios. By examining sample sizes, Mendel showed that his crosses behaved reproducibly according to the laws of probability, and that the traits were inherited as independent events.

Two rules in probability can be used to find the expected proportions of offspring of different traits from different crosses. To find the probability of two or more independent events occurring together, apply the product rule and multiply the probabilities of the individual events. The use of the word “and” suggests the appropriate application of the product rule. To find the probability of two or more events occurring in combination, apply the sum rule and add their individual probabilities together. The use of the word “or” suggests the appropriate application of the sum rule.

Review Questions

Exercise:

Problem:

Mendel performed hybridizations by transferring pollen from the _____ of the male plant to the female ova.

- a. anther
- b. pistil
- c. stigma
- d. seed

Solution:

A

Exercise:

Problem:

Which is one of the seven characteristics that Mendel observed in pea plants?

- a. flower size

- b. seed texture
- c. leaf shape
- d. stem color

Solution:

B

Exercise:

Problem:

Imagine you are performing a cross involving seed color in garden pea plants. What F_1 offspring would you expect if you cross true-breeding parents with green seeds and yellow seeds? Yellow seed color is dominant over green.

- a. 100 percent yellow-green seeds
- b. 100 percent yellow seeds
- c. 50 percent yellow, 50 percent green seeds
- d. 25 percent green, 75 percent yellow seeds

Solution:

B

Exercise:

Problem:

Consider a cross to investigate the pea pod texture trait, involving constricted or inflated pods. Mendel found that the traits behave according to a dominant/recessive pattern in which inflated pods were dominant. If you performed this cross and obtained 650 inflated-pod plants in the F_2 generation, approximately how many constricted-pod plants would you expect to have?

- a. 600
- b. 165
- c. 217
- d. 468

Solution:

C

Free Response

Exercise:

Problem:

Describe one of the reasons why the garden pea was an excellent choice of model system for studying inheritance.

Solution:

The garden pea is sessile and has flowers that close tightly during self-pollination. These features help to prevent accidental or unintentional fertilizations that could have diminished the accuracy of Mendel's data.

Exercise:

Problem:

How would you perform a reciprocal cross for the characteristic of stem height in the garden pea?

Solution:

Two sets of P_0 parents would be used. In the first cross, pollen would be transferred from a true-breeding tall plant to the stigma of a true-breeding dwarf plant. In the second cross, pollen would be transferred from a true-breeding dwarf plant to the stigma of a true-breeding tall plant. For each cross, F_1 and F_2 offspring would be analyzed to determine if offspring traits were affected according to which parent donated each trait.

Glossary

blending theory of inheritance

hypothetical inheritance pattern in which parental traits are blended together in the offspring to produce an intermediate physical appearance

continuous variation

inheritance pattern in which a character shows a range of trait values with small gradations rather than large gaps between them

discontinuous variation

inheritance pattern in which traits are distinct and are transmitted independently of one another

dominant

trait which confers the same physical appearance whether an individual has two copies of the trait or one copy of the dominant trait and one copy of the recessive trait

F_1

first filial generation in a cross; the offspring of the parental generation

F_2

second filial generation produced when F_1 individuals are self-crossed or fertilized with each other

hybridization

process of mating two individuals that differ with the goal of achieving a certain characteristic in their offspring

model system

species or biological system used to study a specific biological phenomenon to be applied to other different species

P_0

parental generation in a cross

product rule

probability of two independent events occurring simultaneously can be calculated by multiplying the individual probabilities of each event occurring alone

recessive

trait that appears “latent” or non-expressed when the individual also carries a dominant trait for that same characteristic; when present as two identical copies, the recessive trait is expressed

reciprocal cross

paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross

sum rule

probability of the occurrence of at least one of two mutually exclusive events is the sum of their individual probabilities

trait

variation in the physical appearance of a heritable characteristic

Characteristics and Traits

By the end of this section, you will be able to:

- Explain the relationship between genotypes and phenotypes in dominant and recessive gene systems
- Develop a Punnett square to calculate the expected proportions of genotypes and phenotypes in a monohybrid cross
- Explain the purpose and methods of a test cross
- Identify non-Mendelian inheritance patterns such as incomplete dominance, codominance, recessive lethals, multiple alleles, and sex linkage

The seven characteristics that Mendel evaluated in his pea plants were each expressed as one of two versions, or traits. The physical expression of characteristics is accomplished through the expression of genes carried on chromosomes. The genetic makeup of peas consists of two similar or homologous copies of each chromosome, one from each parent. Each pair of homologous chromosomes has the same linear order of genes. In other words, peas are diploid organisms in that they have two copies of each chromosome. The same is true for many other plants and for virtually all animals. Diploid organisms utilize meiosis to produce haploid gametes, which contain one copy of each homologous chromosome that unite at fertilization to create a diploid zygote.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. Gene variants that arise by mutation and exist at the same relative locations on homologous chromosomes are called **alleles**. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its **phenotype**. An organism's underlying

genetic makeup, consisting of both physically visible and non-expressed alleles, is called its **genotype**. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. When true-breeding plants in which one parent had yellow pods and one had green pods were cross-fertilized, all of the F_1 hybrid offspring had yellow pods. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow pods. However, we know that the allele donated by the parent with green pods was not simply lost because it reappeared in some of the F_2 offspring. Therefore, the F_1 plants must have been genotypically different from the parent with yellow pods.

The P_1 plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are **homozygous** at a given gene, or locus, have two identical alleles for that gene on their homologous chromosomes. Mendel's parental pea plants always bred true because both of the gametes produced carried the same trait. When P_1 plants with contrasting traits were cross-fertilized, all of the offspring were **heterozygous** for the contrasting trait, meaning that their genotype reflected that they had different alleles for the gene being examined.

Dominant and Recessive Alleles

Our discussion of homozygous and heterozygous organisms brings us to why the F_1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant, and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosome pairs. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype). The recessive allele will only be observed in homozygous recessive individuals ([link](#)).

Human Inheritance in Dominant and Recessive Patterns	
Dominant Traits	Recessive Traits
Achondroplasia	Albinism
Brachydactyly	Cystic fibrosis
Huntington's disease	Duchenne muscular dystrophy
Marfan syndrome	Galactosemia
Neurofibromatosis	Phenylketonuria
Widow's peak	Sickle-cell anemia
Wooly hair	Tay-Sachs disease

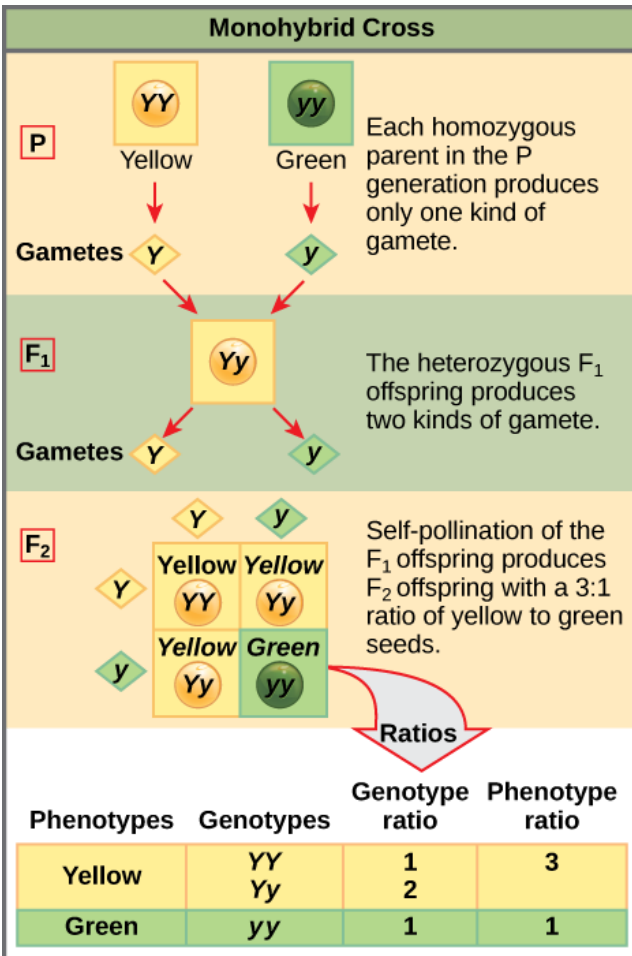
Several conventions exist for referring to genes and alleles. For the purposes of this chapter, we will abbreviate genes using the first letter of the gene's corresponding dominant trait. For example, violet is the dominant trait for a pea plant's flower color, so the flower-color gene would be abbreviated as *V* (note that it is customary to italicize gene designations). Furthermore, we will use uppercase and lowercase letters to represent dominant and recessive alleles, respectively. Therefore, we would refer to the genotype of a homozygous dominant pea plant with violet flowers as *VV*, a homozygous recessive pea plant with white flowers as *vv*, and a heterozygous pea plant with violet flowers as *Vv*.

The Punnett Square Approach for a Monohybrid Cross

When fertilization occurs between two true-breeding parents that differ in only one characteristic, the process is called a **monohybrid** cross, and the resulting offspring are monohybrids. Mendel performed seven monohybrid crosses involving contrasting traits for each characteristic. On the basis of his results in F_1 and F_2 generations, Mendel postulated that each parent in

the monohybrid cross contributed one of two paired unit factors to each offspring, and every possible combination of unit factors was equally likely.

To demonstrate a monohybrid cross, consider the case of true-breeding pea plants with yellow versus green pea seeds. The dominant seed color is yellow; therefore, the parental genotypes were YY for the plants with yellow seeds and yy for the plants with green seeds, respectively. A **Punnett square**, devised by the British geneticist Reginald Punnett, can be drawn that applies the rules of probability to predict the possible outcomes of a genetic cross or mating and their expected frequencies. To prepare a Punnett square, all possible combinations of the parental alleles are listed along the top (for one parent) and side (for the other parent) of a grid, representing their meiotic segregation into haploid gametes. Then the combinations of egg and sperm are made in the boxes in the table to show which alleles are combining. Each box then represents the diploid genotype of a zygote, or fertilized egg, that could result from this mating. Because each possibility is equally likely, genotypic ratios can be determined from a Punnett square. If the pattern of inheritance (dominant or recessive) is known, the phenotypic ratios can be inferred as well. For a monohybrid cross of two true-breeding parents, each parent contributes one type of allele. In this case, only one genotype is possible. All offspring are Yy and have yellow seeds ([\[link\]](#)).



In the P generation, pea plants that are true-breeding for the dominant yellow phenotype are crossed with plants with the recessive green phenotype. This cross produces F₁ heterozygotes with a yellow phenotype. Punnett square analysis can be used to predict the genotypes of the F₂ generation.

A self-cross of one of the Yy heterozygous offspring can be represented in a 2 × 2 Punnett square because each parent can donate one of two different alleles. Therefore, the offspring can potentially have one of four allele

combinations: YY, Yy, yY, or yy ([link](#)). Notice that there are two ways to obtain the Yy genotype: a Y from the egg and a y from the sperm, or a y from the egg and a Y from the sperm. Both of these possibilities must be counted. Recall that Mendel's pea-plant characteristics behaved in the same way in reciprocal crosses. Therefore, the two possible heterozygous combinations produce offspring that are genotypically and phenotypically identical despite their dominant and recessive alleles deriving from different parents. They are grouped together. Because fertilization is a random event, we expect each combination to be equally likely and for the offspring to exhibit a ratio of YY:Yy:yy genotypes of 1:2:1 ([link](#)). Furthermore, because the YY and Yy offspring have yellow seeds and are phenotypically identical, applying the sum rule of probability, we expect the offspring to exhibit a phenotypic ratio of 3 yellow:1 green. Indeed, working with large sample sizes, Mendel observed approximately this ratio in every F₂ generation resulting from crosses for individual traits.

Mendel validated these results by performing an F₃ cross in which he self-crossed the dominant- and recessive-expressing F₂ plants. When he self-crossed the plants expressing green seeds, all of the offspring had green seeds, confirming that all green seeds had homozygous genotypes of yy. When he self-crossed the F₂ plants expressing yellow seeds, he found that one-third of the plants bred true, and two-thirds of the plants segregated at a 3:1 ratio of yellow:green seeds. In this case, the true-breeding plants had homozygous (YY) genotypes, whereas the segregating plants corresponded to the heterozygous (Yy) genotype. When these plants self-fertilized, the outcome was just like the F₁ self-fertilizing cross.

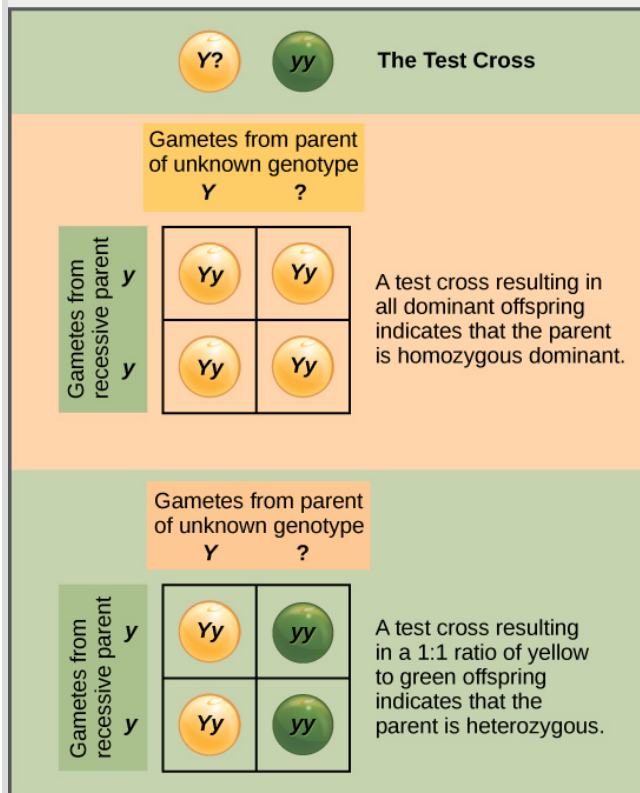
The Test Cross Distinguishes the Dominant Phenotype

Beyond predicting the offspring of a cross between known homozygous or heterozygous parents, Mendel also developed a way to determine whether an organism that expressed a dominant trait was a heterozygote or a homozygote. Called the **test cross**, this technique is still used by plant and animal breeders. In a test cross, the dominant-expressing organism is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then

all F_1 offspring will be heterozygotes expressing the dominant trait ([link](#)). Alternatively, if the dominant expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes ([link](#)). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

Note:

Art Connection



A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

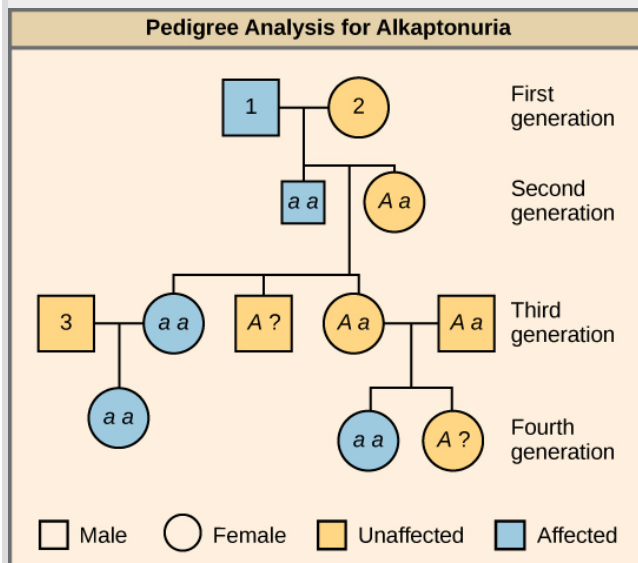
In pea plants, round peas (R) are dominant to wrinkled peas (r). You do a test cross between a pea plant with wrinkled peas (genotype rr) and a plant of unknown genotype that has round peas. You end up with three plants, all

which have round peas. From this data, can you tell if the round pea parent plant is homozygous dominant or heterozygous? If the round pea parent plant is heterozygous, what is the probability that a random sample of 3 progeny peas will all be round?

Many human diseases are genetically inherited. A healthy person in a family in which some members suffer from a recessive genetic disorder may want to know if he or she has the disease-causing gene and what risk exists of passing the disorder on to his or her offspring. Of course, doing a test cross in humans is unethical and impractical. Instead, geneticists use **pedigree analysis** to study the inheritance pattern of human genetic diseases ([link](#)).

Note:

Art Connection



Alkaptonuria is a recessive genetic disorder in which two amino acids, phenylalanine and tyrosine, are not properly metabolized. Affected individuals may have

darkened skin and brown urine, and may suffer joint damage and other complications. In this pedigree, individuals with the disorder are indicated in blue and have the genotype aa . Unaffected individuals are indicated in yellow and have the genotype AA or Aa .

Note that it is often possible to determine a person's genotype from the genotype of their offspring. For example, if neither parent has the disorder but their child does, they must be heterozygous. Two individuals on the pedigree have an unaffected phenotype but unknown genotype.

Because they do not have the disorder, they must have at least one normal allele, so their genotype gets the “A?” designation.

What are the genotypes of the individuals labeled 1, 2 and 3?

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: (1) two “units” or alleles exist for every gene; (2) alleles maintain their integrity in each generation (no blending); and (3) in the presence of the dominant allele, the recessive allele is hidden and makes no contribution to the phenotype. Therefore, recessive alleles can be “carried” and not expressed by individuals. Such heterozygous individuals are sometimes referred to as “carriers.” Further genetic studies in other plants and animals have shown

that much more complexity exists, but that the fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism. If Mendel had chosen an experimental system that exhibited these genetic complexities, it's possible that he would not have understood what his results meant.

Incomplete Dominance

Mendel's results, that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* ([\[link\]](#)), a cross between a homozygous parent with white flowers ($C^W C^W$) and a homozygous parent with red flowers ($C^R C^R$) will produce offspring with pink flowers ($C^R C^W$). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as **incomplete dominance**, denoting the expression of two contrasting alleles such that the individual displays an intermediate phenotype. The allele for red flowers is incompletely dominant over the allele for white flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 $C^R C^R$:2 $C^R C^W$:1 $C^W C^W$, and the phenotypic ratio would be 1:2:1 for red:pink:white.



These pink flowers of a heterozygote snapdragon result from incomplete dominance.

(credit:
“storebukkebruse”/Flickr)

Codominance





A variation on incomplete dominance is **codominance**, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance is the MN blood groups of humans. The M and N alleles are expressed in the form of an M or N antigen present on the surface of red blood cells. Homozygotes ($L^M L^M$ and $L^N L^N$) express either the M or the N allele, and heterozygotes ($L^M L^N$) express both alleles equally. In a self-cross between heterozygotes

expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies.

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level such that many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype among wild animals as the **wild type** (often abbreviated “+”); this is considered the standard or norm. All other phenotypes or genotypes are considered **variants** of this standard, meaning that they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is coat color in rabbits ([link](#)). Here, four alleles exist for the *c* gene. The wild-type version, C^+C^+ , is expressed as brown fur. The chinchilla phenotype, $c^{ch}c^{ch}$, is expressed as black-tipped white fur. The Himalayan phenotype, c^hc^h , has black fur on the extremities and white fur elsewhere. Finally, the albino, or “colorless” phenotype, cc , is expressed as white fur. In cases of multiple alleles, dominance hierarchies can exist. In this case, the wild-type allele is dominant over all the others, chinchilla is incompletely dominant over Himalayan and albino, and Himalayan is dominant over albino. This hierarchy, or allelic series, was revealed by observing the phenotypes of each possible heterozygote offspring.

Allele			
C	c ^{ch}	c ^h	c
Genotype			
CC	c ^{ch} c ^{ch}	c ^h c ^h	cc
Phenotype			
WILD TYPE: Brown fur	CHINCHILLA: Black-tipped white fur	HIMALAYAN: White fur with black paws, nose, ears, tail	ALBINO: White fur
			

Four different alleles exist for the rabbit coat color (C) gene.

The complete dominance of a wild-type phenotype over all other mutants often occurs as an effect of “dosage” of a specific gene product, such that the wild-type allele supplies the correct amount of gene product whereas the mutant alleles cannot. For the allelic series in rabbits, the wild-type allele may supply a given dosage of fur pigment, whereas the mutants supply a lesser dosage or none at all. Interestingly, the Himalayan phenotype is the result of an allele that produces a temperature-sensitive gene product that only produces pigment in the cooler extremities of the rabbit’s body.

Alternatively, one mutant allele can be dominant over all other phenotypes, including the wild type. This may occur when the mutant allele somehow interferes with the genetic message so that even a heterozygote with one wild-type allele copy expresses the mutant phenotype. One way in which the mutant allele can interfere is by enhancing the function of the wild-type gene product or changing its distribution in the body. One example of this is the *Antennapedia* mutation in *Drosophila* ([link](#)). In this case, the mutant allele expands the distribution of the gene product, and as a result, the

Antennapedia heterozygote develops legs on its head where its antennae should be.



As seen in comparing the wild-type *Drosophila* (left) and the *Antennapedia* mutant (right), the *Antennapedia* mutant has legs on its head in place of antennae.

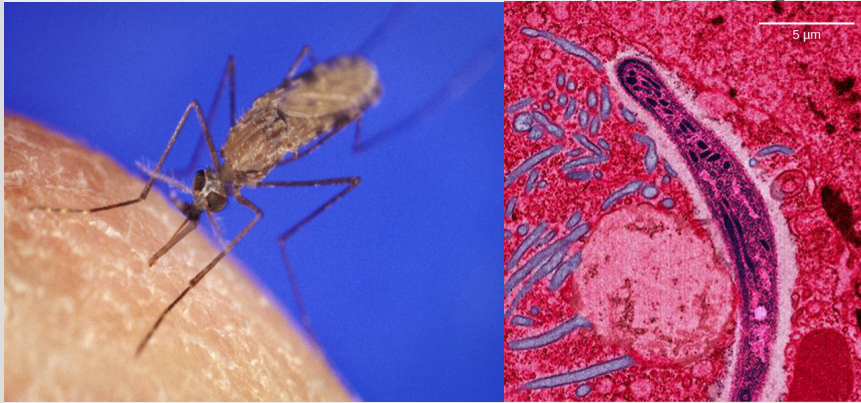
Note:

Evolution Connection

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae* ([\[link\]a](#)), and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly ([\[link\]b](#)). When promptly and correctly treated, *P. falciparum* malaria has a mortality

rate of 0.1 percent. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most effective malarial treatments can vary by geographic region.



The (a) *Anopheles gambiae*, or African malaria mosquito, acts as a vector in the transmission to humans of the malaria-causing parasite (b) *Plasmodium falciparum*, here visualized using false-color transmission electron microscopy. (credit a: James D. Gathany; credit b: Ute Frevert; false color by Margaret Shear; scale-bar data from Matt Russell)

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it is infectious to humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait. In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that occurs because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions where this drug is widely used as an over-the-counter malaria remedy. As is common with pathogens that multiply to large

numbers within an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden. [\[footnote\]](#)

Sumiti Vinayak, et al., “Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*,” *Public Library of Science Pathogens* 6, no. 3 (2010): e1000830, doi:10.1371/journal.ppat.1000830.

X-Linked Traits

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes. The sex chromosomes are one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or **autosomes**. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains many fewer genes. When a gene being examined is present on the X chromosome, but not on the Y chromosome, it is said to be **X-linked**.

Eye color in *Drosophila* was one of the first X-linked traits to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies, the wild-type eye color is red (X^W) and it is dominant to white eye color (X^w) ([\[link\]](#)). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. Males are said to be **hemizygous**, because they have only one allele for any X-linked characteristic. Hemizyosity makes the descriptions of dominance and recessiveness irrelevant for XY males. *Drosophila* males lack a second allele copy on the Y chromosome; that is, their genotype can only be X^WY or X^wY . In contrast, females have two allele copies of this gene and can be X^WX^W , X^WX^w , or X^wX^w .



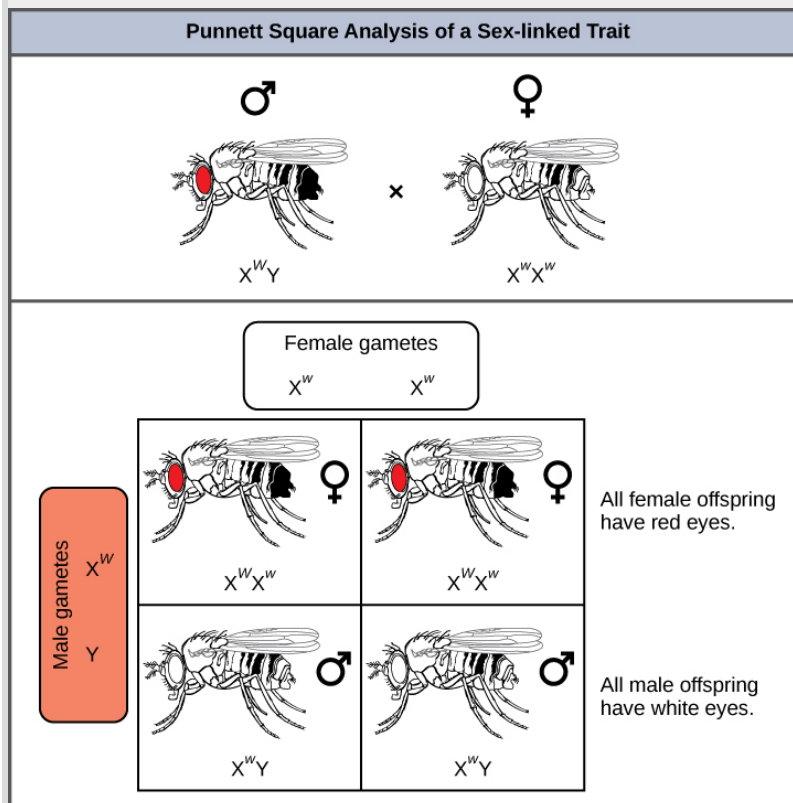
In *Drosophila*, several genes determine eye color. The genes for white and vermilion eye colors are located on the X chromosome. Others are located on the autosomes. Clockwise from top left are brown, cinnabar, sepia, vermilion, white, and red. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P_1 generation. With regard to *Drosophila* eye color, when the P_1 male expresses the white-eye phenotype and the female is homozygous red-eyed, all members of the F_1 generation exhibit red eyes ([link](#)). The F_1 females

are heterozygous ($X^W X^w$), and the males are all $X^W Y$, having received their X chromosome from the homozygous dominant P_1 female and their Y chromosome from the P_1 male. A subsequent cross between the $X^W X^w$ female and the $X^W Y$ male would produce only red-eyed females (with $X^W X^W$ or $X^W X^w$ genotypes) and both red- and white-eyed males (with $X^W Y$ or $X^w Y$ genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females ($X^W X^w$) and only white-eyed males ($X^w Y$). Half of the F_2 females would be red-eyed ($X^W X^W$) and half would be white-eyed ($X^w X^w$). Similarly, half of the F_2 males would be red-eyed ($X^W Y$) and half would be white-eyed ($X^w Y$).

Note:

Art Connection



Punnett square analysis is used to determine the ratio of offspring from a cross between a

red-eyed male fruit fly and a white-eyed
female fruit fly.

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

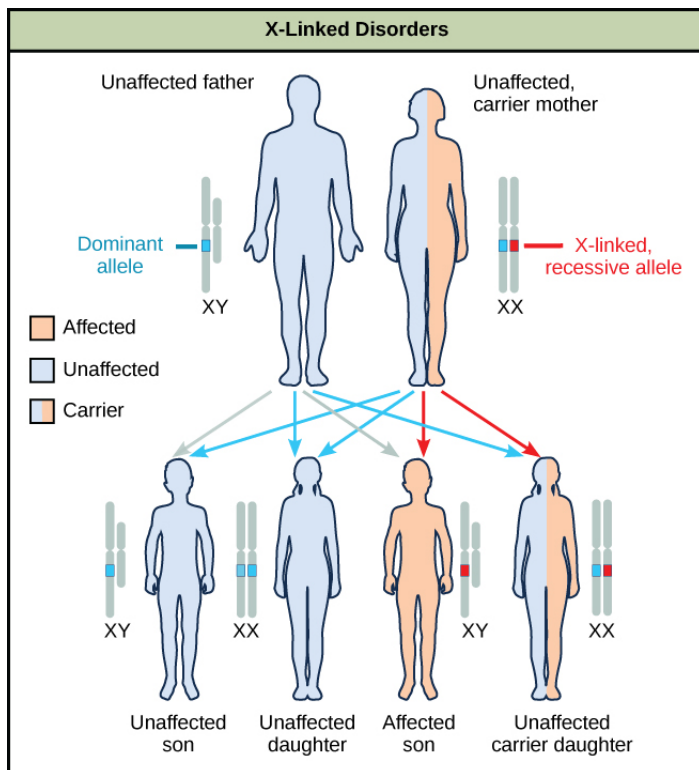
Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the trait on to 100 percent of her offspring. Her male offspring are, therefore, destined to express the trait, as they will inherit their father's Y chromosome. In humans, the alleles for certain conditions (some forms of color blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, recessive X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in which they are hemizygous.

Human Sex-linked Disorders

Sex-linkage studies in Morgan's laboratory provided the fundamentals for understanding X-linked recessive disorders in humans, which include red-green color blindness, and Types A and B hemophilia. Because human males need to inherit only one recessive mutant X allele to be affected, X-linked disorders are disproportionately observed in males. Females must inherit recessive X-linked alleles from both of their parents in order to express the trait. When they inherit one recessive X-linked mutant allele and one dominant X-linked wild-type allele, they are carriers of the trait and are typically unaffected. Carrier females can manifest mild forms of the trait

due to the inactivation of the dominant allele located on one of the X chromosomes. However, female carriers can contribute the trait to their sons, resulting in the son exhibiting the trait, or they can contribute the recessive allele to their daughters, resulting in the daughters being carriers of the trait ([link](#)). Although some Y-linked recessive disorders exist, typically they are associated with infertility in males and are therefore not transmitted to subsequent generations.



The son of a woman who is a carrier of a recessive X-linked disorder will have a 50 percent chance of being affected. A daughter will not be affected, but she will have a 50 percent chance of being a carrier like her mother.

Note:

Link to Learning



Watch this video to learn more about sex-linked traits.

https://www.openstaxcollege.org/l/sex-linked_traits

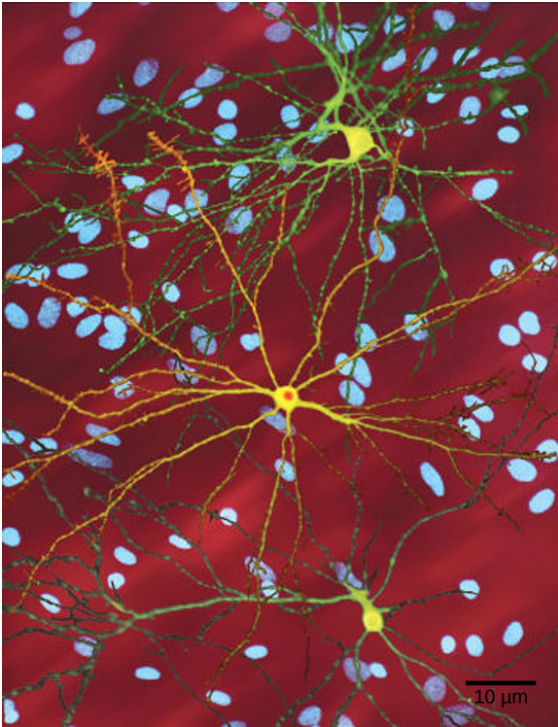
Lethality

A large proportion of genes in an individual's genome are essential for survival. Occasionally, a nonfunctional allele for an essential gene can arise by mutation and be transmitted in a population as long as individuals with this allele also have a wild-type, functional copy. The wild-type allele functions at a capacity sufficient to sustain life and is therefore considered to be dominant over the nonfunctional allele. However, consider two heterozygous parents that have a genotype of wild-type/nonfunctional mutant for a hypothetical essential gene. In one quarter of their offspring, we would expect to observe individuals that are homozygous recessive for the nonfunctional allele. Because the gene is essential, these individuals might fail to develop past fertilization, die *in utero*, or die later in life, depending on what life stage requires this gene. An inheritance pattern in which an allele is only lethal in the homozygous form and in which the heterozygote may be normal or have some altered non-lethal phenotype is referred to as **recessive lethal**.

For crosses between heterozygous individuals with a recessive lethal allele that causes death before birth when homozygous, only wild-type homozygotes and heterozygotes would be observed. The genotypic ratio would therefore be 2:1. In other instances, the recessive lethal allele might also exhibit a dominant (but not lethal) phenotype in the heterozygote. For

instance, the recessive lethal *Curly* allele in *Drosophila* affects wing shape in the heterozygote form but is lethal in the homozygote.

A single copy of the wild-type allele is not always sufficient for normal functioning or even survival. The **dominant lethal** inheritance pattern is one in which an allele is lethal both in the homozygote and the heterozygote; this allele can only be transmitted if the lethality phenotype occurs after reproductive age. Individuals with mutations that result in dominant lethal alleles fail to survive even in the heterozygote form. Dominant lethal alleles are very rare because, as you might expect, the allele only lasts one generation and is not transmitted. However, just as the recessive lethal allele might not immediately manifest the phenotype of death, dominant lethal alleles also might not be expressed until adulthood. Once the individual reaches reproductive age, the allele may be unknowingly passed on, resulting in a delayed death in both generations. An example of this in humans is Huntington's disease, in which the nervous system gradually wastes away ([\[link\]](#)). People who are heterozygous for the dominant Huntington allele (*Hh*) will inevitably develop the fatal disease. However, the onset of Huntington's disease may not occur until age 40, at which point the afflicted persons may have already passed the allele to 50 percent of their offspring.



The neuron in the center of this micrograph (yellow) has nuclear inclusions characteristic of Huntington's disease (orange area in the center of the neuron).

Huntington's disease occurs when an abnormal dominant allele for the Huntington gene is present. (credit: Dr. Steven Finkbeiner, Gladstone Institute of Neurological Disease, The Taube-Koret Center for Huntington's Disease Research, and the University of California San Francisco/Wikimedia)

Section Summary

When true-breeding or homozygous individuals that differ for a certain trait are crossed, all of the offspring will be heterozygotes for that trait. If the traits are inherited as dominant and recessive, the F_1 offspring will all exhibit the same phenotype as the parent homozygous for the dominant trait. If these heterozygous offspring are self-crossed, the resulting F_2 offspring will be equally likely to inherit gametes carrying the dominant or recessive trait, giving rise to offspring of which one quarter are homozygous dominant, half are heterozygous, and one quarter are homozygous recessive. Because homozygous dominant and heterozygous individuals are phenotypically identical, the observed traits in the F_2 offspring will exhibit a ratio of three dominant to one recessive.

Alleles do not always behave in dominant and recessive patterns. Incomplete dominance describes situations in which the heterozygote exhibits a phenotype that is intermediate between the homozygous phenotypes. Codominance describes the simultaneous expression of both of the alleles in the heterozygote. Although diploid organisms can only have two alleles for any given gene, it is common for more than two alleles of a gene to exist in a population. In humans, as in many animals and some plants, females have two X chromosomes and males have one X and one Y chromosome. Genes that are present on the X but not the Y chromosome are said to be X-linked, such that males only inherit one allele for the gene, and females inherit two. Finally, some alleles can be lethal. Recessive lethal alleles are only lethal in homozygotes, but dominant lethal alleles are fatal in heterozygotes as well.

Art Connections

Exercise:

Problem:

[\[link\]](#) In pea plants, round peas (R) are dominant to wrinkled peas (r). You do a test cross between a pea plant with wrinkled peas (genotype rr) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas. From this data, can you tell if the round pea parent plant is homozygous dominant or heterozygous? If the round pea parent plant is heterozygous, what is the probability that a random sample of 3 progeny peas will all be round?

Solution:

[\[link\]](#) You cannot be sure if the plant is homozygous or heterozygous as the data set is too small: by random chance, all three plants might have acquired only the dominant gene even if the recessive one is present. If the round pea parent is heterozygous, there is a one-eighth probability that a random sample of three progeny peas will all be round.

Exercise:**Problem:**

[\[link\]](#) What are the genotypes of the individuals labeled 1, 2 and 3?

Solution:

[\[link\]](#) Individual 1 has the genotype aa . Individual 2 has the genotype Aa . Individual 3 has the genotype Aa .

Exercise:**Problem:**

[\[link\]](#) What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Solution:

[\[link\]](#) Half of the female offspring would be heterozygous ($X^W X^w$) with red eyes, and half would be homozygous recessive ($X^w X^w$) with white eyes. Half of the male offspring would be hemizygous dominant ($X^W Y$) with red eyes, and half would be hemizygous recessive ($X^w Y$) with white eyes.

Review Questions

Exercise:

Problem:

The observable traits expressed by an organism are described as its _____.

- a. phenotype
- b. genotype
- c. alleles
- d. zygote

Solution:

A

Exercise:

Problem:

A recessive trait will be observed in individuals that are _____ for that trait.

- a. heterozygous
- b. homozygous or heterozygous
- c. homozygous
- d. diploid

Solution:

C

Exercise:

Problem:

If black and white true-breeding mice are mated and the result is all gray offspring, what inheritance pattern would this be indicative of?

- a. dominance
- b. codominance
- c. multiple alleles
- d. incomplete dominance

Solution:

D

Exercise:

Problem:

The ABO blood groups in humans are expressed as the I^A , I^B , and i alleles. The I^A allele encodes the A blood group antigen, I^B encodes B, and i encodes O. Both A and B are dominant to O. If a heterozygous blood type A parent ($I^A i$) and a heterozygous blood type B parent ($I^B i$) mate, one quarter of their offspring will have AB blood type ($I^A I^B$) in which both antigens are expressed equally. Therefore, ABO blood groups are an example of:

- a. multiple alleles and incomplete dominance
- b. codominance and incomplete dominance
- c. incomplete dominance only
- d. multiple alleles and codominance

Solution:

D

Exercise:**Problem:**

In a mating between two individuals that are heterozygous for a recessive lethal allele that is expressed *in utero*, what genotypic ratio (homozygous dominant:heterozygous:homozygous recessive) would you expect to observe in the offspring?

- a. 1:2:1
- b. 3:1:1
- c. 1:2:0
- d. 0:2:1

Solution:

C

Free Response**Exercise:****Problem:**

The gene for flower position in pea plants exists as axial or terminal alleles. Given that axial is dominant to terminal, list all of the possible F_1 and F_2 genotypes and phenotypes from a cross involving parents that are homozygous for each trait. Express genotypes with conventional genetic abbreviations.

Solution:

Because axial is dominant, the gene would be designated as A . F_1 would be all heterozygous Aa with axial phenotype. F_2 would have possible genotypes of AA , Aa , and aa ; these would correspond to axial, axial, and terminal phenotypes, respectively.

Exercise:

Problem:

Use a Punnett square to predict the offspring in a cross between a dwarf pea plant (homozygous recessive) and a tall pea plant (heterozygous). What is the phenotypic ratio of the offspring?

Solution:

The Punnett square would be 2×2 and will have T and T along the top, and T and t along the left side. Clockwise from the top left, the genotypes listed within the boxes will be Tt , Tt , tt , and tt . The phenotypic ratio will be 1 tall:1 dwarf.

Exercise:

Problem: Can a human male be a carrier of red-green color blindness?

Solution:

No, males can only express color blindness. They cannot carry it because an individual needs two X chromosomes to be a carrier.

Glossary

allele

gene variations that arise by mutation and exist at the same relative locations on homologous chromosomes

autosomes

any of the non-sex chromosomes

codominance

in a heterozygote, complete and simultaneous expression of both alleles for the same characteristic

dominant lethal

inheritance pattern in which an allele is lethal both in the homozygote and the heterozygote; this allele can only be transmitted if the lethality phenotype occurs after reproductive age

genotype

underlying genetic makeup, consisting of both physically visible and non-expressed alleles, of an organism

hemizygous

presence of only one allele for a characteristic, as in X-linkage; hemizyosity makes descriptions of dominance and recessiveness irrelevant

heterozygous

having two different alleles for a given gene on the homologous chromosome

homozygous

having two identical alleles for a given gene on the homologous chromosome

incomplete dominance

in a heterozygote, expression of two contrasting alleles such that the individual displays an intermediate phenotype

monohybrid

result of a cross between two true-breeding parents that express different traits for only one characteristic

phenotype

observable traits expressed by an organism

Punnett square

visual representation of a cross between two individuals in which the gametes of each individual are denoted along the top and side of a grid, respectively, and the possible zygotic genotypes are recombined at each box in the grid

recessive lethal

inheritance pattern in which an allele is only lethal in the homozygous form; the heterozygote may be normal or have some altered, non-lethal phenotype

sex-linked

any gene on a sex chromosome

test cross

cross between a dominant expressing individual with an unknown genotype and a homozygous recessive individual; the offspring phenotypes indicate whether the unknown parent is heterozygous or homozygous for the dominant trait

X-linked

gene present on the X, but not the Y chromosome

Laws of Inheritance

By the end of this section, you will be able to:

- Explain Mendel's law of segregation and independent assortment in terms of genetics and the events of meiosis
- Use the forked-line method and the probability rules to calculate the probability of genotypes and phenotypes from multiple gene crosses
- Explain the effect of linkage and recombination on gamete genotypes
- Explain the phenotypic outcomes of epistatic effects between genes

Mendel generalized the results of his pea-plant experiments into four postulates, some of which are sometimes called “laws,” that describe the basis of dominant and recessive inheritance in diploid organisms. As you have learned, more complex extensions of Mendelism exist that do not exhibit the same F_2 phenotypic ratios (3:1). Nevertheless, these laws summarize the basics of classical genetics.

Pairs of Unit Factors, or Genes

Mendel proposed first that paired unit factors of heredity were transmitted faithfully from generation to generation by the dissociation and reassociation of paired factors during gametogenesis and fertilization, respectively. After he crossed peas with contrasting traits and found that the recessive trait resurfaced in the F_2 generation, Mendel deduced that hereditary factors must be inherited as discrete units. This finding contradicted the belief at that time that parental traits were blended in the offspring.

Alleles Can Be Dominant or Recessive

Mendel's **law of dominance** states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. Rather than both alleles contributing to a phenotype, the dominant allele will be expressed exclusively. The recessive allele will remain “latent” but will be transmitted to offspring by the same manner in which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that

have two copies of this allele ([link](#)), and these offspring will breed true when self-crossed.

Since Mendel's experiments with pea plants, other researchers have found that the law of dominance does not always hold true. Instead, several different patterns of inheritance have been found to exist.



The child in the photo expresses albinism, a recessive trait.

Equal Segregation of Alleles

Observing that true-breeding pea plants with contrasting traits gave rise to F_1 generations that all expressed the dominant trait and F_2 generations that expressed the dominant and recessive traits in a 3:1 ratio, Mendel proposed the **law of segregation**. This law states that paired unit factors (genes) must segregate equally into gametes such that offspring have an equal likelihood

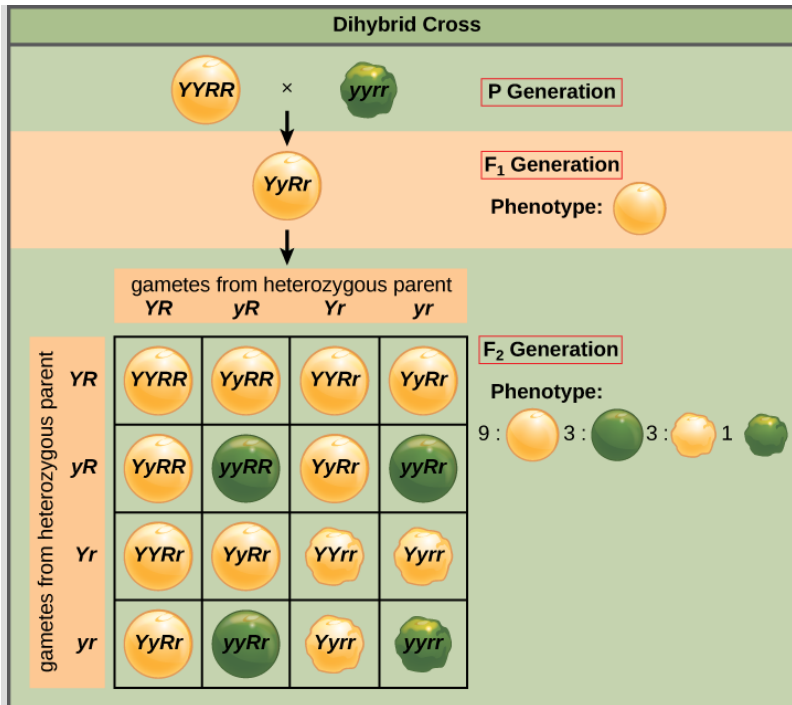
of inheriting either factor. For the F_2 generation of a monohybrid cross, the following three possible combinations of genotypes could result: homozygous dominant, heterozygous, or homozygous recessive. Because heterozygotes could arise from two different pathways (receiving one dominant and one recessive allele from either parent), and because heterozygotes and homozygous dominant individuals are phenotypically identical, the law supports Mendel's observed 3:1 phenotypic ratio. The equal segregation of alleles is the reason we can apply the Punnett square to accurately predict the offspring of parents with known genotypes. The physical basis of Mendel's law of segregation is the first division of meiosis, in which the homologous chromosomes with their different versions of each gene are segregated into daughter nuclei. The role of the meiotic segregation of chromosomes in sexual reproduction was not understood by the scientific community during Mendel's lifetime.

Independent Assortment

Mendel's **law of independent assortment** states that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. The independent assortment of genes can be illustrated by the **dihybrid** cross, a cross between two true-breeding parents that express different traits for two characteristics. Consider the characteristics of seed color and seed texture for two pea plants, one that has green, wrinkled seeds ($yyrr$) and another that has yellow, round seeds ($YYRR$). Because each parent is homozygous, the law of segregation indicates that the gametes for the green/wrinkled plant all are yr , and the gametes for the yellow/round plant are all YR . Therefore, the F_1 generation of offspring all are $YyRr$ ([link](#)).

Note:

Art Connection



This dihybrid cross of pea plants involves the genes for seed color and texture.

In pea plants, purple flowers (P) are dominant to white flowers (p) and yellow peas (Y) are dominant to green peas (y). What are the possible genotypes and phenotypes for a cross between PpYY and ppYy pea plants? How many squares do you need to do a Punnett square analysis of this cross?

For the F₂ generation, the law of segregation requires that each gamete receive either an *R* allele or an *r* allele along with either a *Y* allele or a *y* allele. The law of independent assortment states that a gamete into which an *r* allele sorted would be equally likely to contain either a *Y* allele or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *YyRr* heterozygote is self-crossed, as follows: *YR*, *Yr*, *yR*, and *yr*. Arranging these gametes along the top and left of a 4 × 4 Punnett square ([link](#)) gives us 16 equally likely genotypic combinations. From these genotypes, we infer a phenotypic ratio of 9 round/yellow:3 round/green:3

wrinkled/yellow:1 wrinkled/green ([link](#)). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

Because of independent assortment and dominance, the 9:3:3:1 dihybrid phenotypic ratio can be collapsed into two 3:1 ratios, characteristic of any monohybrid cross that follows a dominant and recessive pattern. Ignoring seed color and considering only seed texture in the above dihybrid cross, we would expect that three quarters of the F_2 generation offspring would be round, and one quarter would be wrinkled. Similarly, isolating only seed color, we would assume that three quarters of the F_2 offspring would be yellow and one quarter would be green. The sorting of alleles for texture and color are independent events, so we can apply the product rule.

Therefore, the proportion of round and yellow F_2 offspring is expected to be $(3/4) \times (3/4) = 9/16$, and the proportion of wrinkled and green offspring is expected to be $(1/4) \times (1/4) = 1/16$. These proportions are identical to those obtained using a Punnett square. Round, green and wrinkled, yellow offspring can also be calculated using the product rule, as each of these genotypes includes one dominant and one recessive phenotype. Therefore, the proportion of each is calculated as $(3/4) \times (1/4) = 3/16$.

The law of independent assortment also indicates that a cross between yellow, wrinkled ($YYrr$) and green, round ($yyRR$) parents would yield the same F_1 and F_2 offspring as in the $YYRR \times yyrr$ cross.

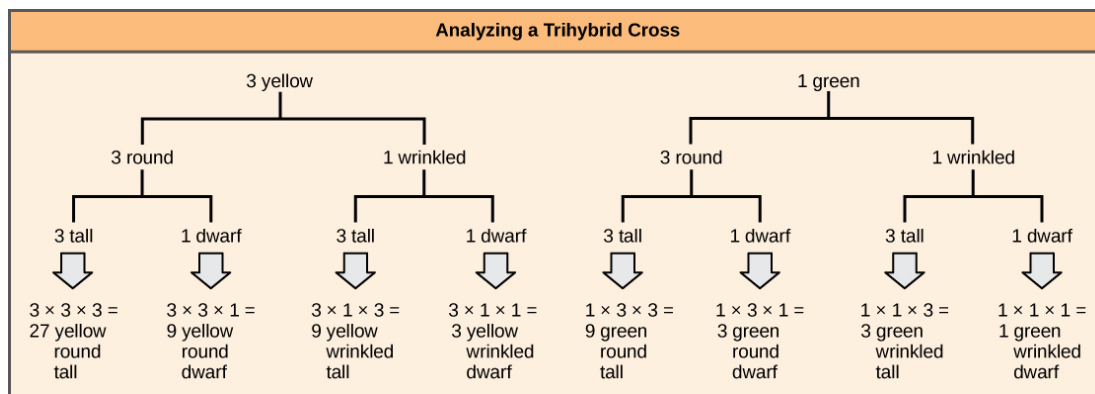
The physical basis for the law of independent assortment also lies in meiosis I, in which the different homologous pairs line up in random orientations. Each gamete can contain any combination of paternal and maternal chromosomes (and therefore the genes on them) because the orientation of tetrads on the metaphase plane is random.

Forked-Line Method

When more than two genes are being considered, the Punnett-square method becomes unwieldy. For instance, examining a cross involving four genes would require a 16×16 grid containing 256 boxes. It would be

extremely cumbersome to manually enter each genotype. For more complex crosses, the forked-line and probability methods are preferred.

To prepare a forked-line diagram for a cross between F_1 heterozygotes resulting from a cross between $AABBCC$ and $aabbcc$ parents, we first create rows equal to the number of genes being considered, and then segregate the alleles in each row on forked lines according to the probabilities for individual monohybrid crosses ([link](#)). We then multiply the values along each forked path to obtain the F_2 offspring probabilities. Note that this process is a diagrammatic version of the product rule. The values along each forked pathway can be multiplied because each gene assorts independently. For a trihybrid cross, the F_2 phenotypic ratio is 27:9:9:9:3:3:3:1.



The forked-line method can be used to analyze a trihybrid cross. Here, the probability for color in the F_2 generation occupies the top row (3 yellow:1 green). The probability for shape occupies the second row (3 round:1 wrinkled), and the probability for height occupies the third row (3 tall:1 dwarf).

The probability for each possible combination of traits is calculated by multiplying the probability for each individual trait. Thus, the probability of F_2 offspring having yellow, round, and tall traits is $3 \times 3 \times 3$, or 27.

Probability Method

While the forked-line method is a diagrammatic approach to keeping track of probabilities in a cross, the probability method gives the proportions of offspring expected to exhibit each phenotype (or genotype) without the added visual assistance. Both methods make use of the product rule and consider the alleles for each gene separately. Earlier, we examined the phenotypic proportions for a trihybrid cross using the forked-line method; now we will use the probability method to examine the genotypic proportions for a cross with even more genes.

For a trihybrid cross, writing out the forked-line method is tedious, albeit not as tedious as using the Punnett-square method. To fully demonstrate the power of the probability method, however, we can consider specific genetic calculations. For instance, for a tetrahybrid cross between individuals that are heterozygotes for all four genes, and in which all four genes are sorting independently and in a dominant and recessive pattern, what proportion of the offspring will be expected to be homozygous recessive for all four alleles? Rather than writing out every possible genotype, we can use the probability method. We know that for each gene, the fraction of homozygous recessive offspring will be $1/4$. Therefore, multiplying this fraction for each of the four genes, $(1/4) \times (1/4) \times (1/4) \times (1/4)$, we determine that $1/256$ of the offspring will be quadruply homozygous recessive.

For the same tetrahybrid cross, what is the expected proportion of offspring that have the dominant phenotype at all four loci? We can answer this question using phenotypic proportions, but let's do it the hard way—using genotypic proportions. The question asks for the proportion of offspring that are 1) homozygous dominant at *A* or heterozygous at *A*, and 2) homozygous at *B* or heterozygous at *B*, and so on. Noting the “or” and “and” in each circumstance makes clear where to apply the sum and product rules. The probability of a homozygous dominant at *A* is $1/4$ and the probability of a heterozygote at *A* is $1/2$. The probability of the homozygote or the heterozygote is $1/4 + 1/2 = 3/4$ using the sum rule. The same probability can be obtained in the same way for each of the other genes, so that the probability of a dominant phenotype at *A* and *B* and *C* and

D is, using the product rule, equal to $\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4}$, or 27/64. If you are ever unsure about how to combine probabilities, returning to the forked-line method should make it clear.

Rules for Multihybrid Fertilization

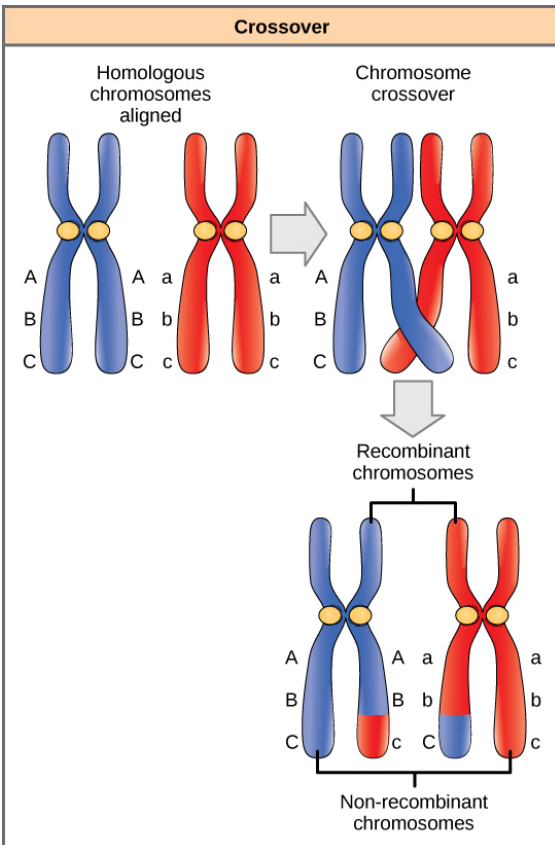
Predicting the genotypes and phenotypes of offspring from given crosses is the best way to test your knowledge of Mendelian genetics. Given a multihybrid cross that obeys independent assortment and follows a dominant and recessive pattern, several generalized rules exist; you can use these rules to check your results as you work through genetics calculations ([\[link\]](#)). To apply these rules, first you must determine *n*, the number of heterozygous gene pairs (the number of genes segregating two alleles each). For example, a cross between *AaBb* and *AaBb* heterozygotes has an *n* of 2. In contrast, a cross between *AABb* and *AABb* has an *n* of 1 because *A* is not heterozygous.

General Rules for Multihybrid Crosses	
General Rule	Number of Heterozygous Gene Pairs
Number of different F ₁ gametes	2 ^{<i>n</i>}
Number of different F ₂ genotypes	3 ^{<i>n</i>}
Given dominant and recessive inheritance, the number of different F ₂ phenotypes	2 ^{<i>n</i>}

Linked Genes Violate the Law of Independent Assortment

Although all of Mendel's pea characteristics behaved according to the law of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on separate non-homologous chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. The segregation of alleles into gametes can be influenced by **linkage**, in which genes that are located physically close to each other on the same chromosome are more likely to be inherited as a pair. However, because of the process of recombination, or "crossover," it is possible for two genes on the same chromosome to behave independently, or as if they are not linked. To understand this, let's consider the biological basis of gene linkage and recombination.

Homologous chromosomes possess the same genes in the same linear order. The alleles may differ on homologous chromosome pairs, but the genes to which they correspond do not. In preparation for the first division of meiosis, homologous chromosomes replicate and synapse. Like genes on the homologs align with each other. At this stage, segments of homologous chromosomes exchange linear segments of genetic material ([link](#)). This process is called recombination, or crossover, and it is a common genetic process. Because the genes are aligned during recombination, the gene order is not altered. Instead, the result of recombination is that maternal and paternal alleles are combined onto the same chromosome. Across a given chromosome, several recombination events may occur, causing extensive shuffling of alleles.



The process of crossover, or recombination, occurs when two homologous chromosomes align during meiosis and exchange a segment of genetic material. Here, the alleles for gene C were exchanged. The result is two recombinant and two non-recombinant chromosomes.

When two genes are located in close proximity on the same chromosome, they are considered linked, and their alleles tend to be transmitted through meiosis together. To exemplify this, imagine a dihybrid cross involving flower color and plant height in which the genes are next to each other on

the chromosome. If one homologous chromosome has alleles for tall plants and red flowers, and the other chromosome has genes for short plants and yellow flowers, then when the gametes are formed, the tall and red alleles will go together into a gamete and the short and yellow alleles will go into other gametes. These are called the parental genotypes because they have been inherited intact from the parents of the individual producing gametes. But unlike if the genes were on different chromosomes, there will be no gametes with tall and yellow alleles and no gametes with short and red alleles. If you create the Punnett square with these gametes, you will see that the classical Mendelian prediction of a 9:3:3:1 outcome of a dihybrid cross would not apply. As the distance between two genes increases, the probability of one or more crossovers between them increases, and the genes behave more like they are on separate chromosomes. Geneticists have used the proportion of recombinant gametes (the ones not like the parents) as a measure of how far apart genes are on a chromosome. Using this information, they have constructed elaborate maps of genes on chromosomes for well-studied organisms, including humans.

Mendel's seminal publication makes no mention of linkage, and many researchers have questioned whether he encountered linkage but chose not to publish those crosses out of concern that they would invalidate his independent assortment postulate. The garden pea has seven chromosomes, and some have suggested that his choice of seven characteristics was not a coincidence. However, even if the genes he examined were not located on separate chromosomes, it is possible that he simply did not observe linkage because of the extensive shuffling effects of recombination.

Note:

Scientific Method Connection

Testing the Hypothesis of Independent Assortment

To better appreciate the amount of labor and ingenuity that went into Mendel's experiments, proceed through one of Mendel's dihybrid crosses.

Question: What will be the offspring of a dihybrid cross?

Background: Consider that pea plants mature in one growing season, and you have access to a large garden in which you can cultivate thousands of

pea plants. There are several true-breeding plants with the following pairs of traits: tall plants with inflated pods, and dwarf plants with constricted pods. Before the plants have matured, you remove the pollen-producing organs from the tall/inflated plants in your crosses to prevent self-fertilization. Upon plant maturation, the plants are manually crossed by transferring pollen from the dwarf/constricted plants to the stigmata of the tall/inflated plants.

Hypothesis: Both trait pairs will sort independently according to Mendelian laws. When the true-breeding parents are crossed, all of the F_1 offspring are tall and have inflated pods, which indicates that the tall and inflated traits are dominant over the dwarf and constricted traits, respectively. A self-cross of the F_1 heterozygotes results in 2,000 F_2 progeny.

Test the hypothesis: Because each trait pair sorts independently, the ratios of tall:dwarf and inflated:constricted are each expected to be 3:1. The tall/dwarf trait pair is called T/t , and the inflated/constricted trait pair is designated I/i . Each member of the F_1 generation therefore has a genotype of $TtIi$. Construct a grid analogous to [\[link\]](#), in which you cross two $TtIi$ individuals. Each individual can donate four combinations of two traits: TI , Ti , tI , or ti , meaning that there are 16 possibilities of offspring genotypes. Because the T and I alleles are dominant, any individual having one or two of those alleles will express the tall or inflated phenotypes, respectively, regardless if they also have a t or i allele. Only individuals that are tt or ii will express the dwarf and constricted alleles, respectively. As shown in [\[link\]](#), you predict that you will observe the following offspring proportions: tall/inflated:tall/constricted:dwarf/inflated:dwarf/constricted in a 9:3:3:1 ratio. Notice from the grid that when considering the tall/dwarf and inflated/constricted trait pairs in isolation, they are each inherited in 3:1 ratios.

		<i>TtIi</i>			
		<i>TI</i>	<i>Ti</i>	<i>tI</i>	<i>ti</i>
<i>TtIi</i>	<i>TI</i>	<i>TTII</i>	<i>TTIi</i>	<i>TtII</i>	<i>TtIi</i>
	<i>Ti</i>	<i>TTIi</i>	<i>TTii</i>	<i>TtIi</i>	<i>Ttii</i>
	<i>tI</i>	<i>TtII</i>	<i>TtIi</i>	<i>ttII</i>	<i>ttIi</i>
	<i>ti</i>	<i>TtIi</i>	<i>Ttii</i>	<i>ttIi</i>	<i>ttii</i>

This figure shows all possible combinations of offspring resulting from a dihybrid cross of pea plants that are heterozygous for the tall/dwarf and inflated/constricted alleles.

Test the hypothesis: You cross the dwarf and tall plants and then self-cross the offspring. For best results, this is repeated with hundreds or even thousands of pea plants. What special precautions should be taken in the crosses and in growing the plants?

Analyze your data: You observe the following plant phenotypes in the F₂ generation: 2706 tall/inflated, 930 tall/constricted, 888 dwarf/inflated, and 300 dwarf/constricted. Reduce these findings to a ratio and determine if they are consistent with Mendelian laws.

Form a conclusion: Were the results close to the expected 9:3:3:1 phenotypic ratio? Do the results support the prediction? What might be observed if far fewer plants were used, given that alleles segregate randomly into gametes? Try to imagine growing that many pea plants, and consider the potential for experimental error. For instance, what would happen if it was extremely windy one day?

Epistasis

Mendel's studies in pea plants implied that the sum of an individual's phenotype was controlled by genes (or as he called them, unit factors), such that every characteristic was distinctly and completely controlled by a single gene. In fact, single observable characteristics are almost always under the influence of multiple genes (each with two or more alleles) acting in unison. For example, at least eight genes contribute to eye color in humans.

Note:

Link to Learning



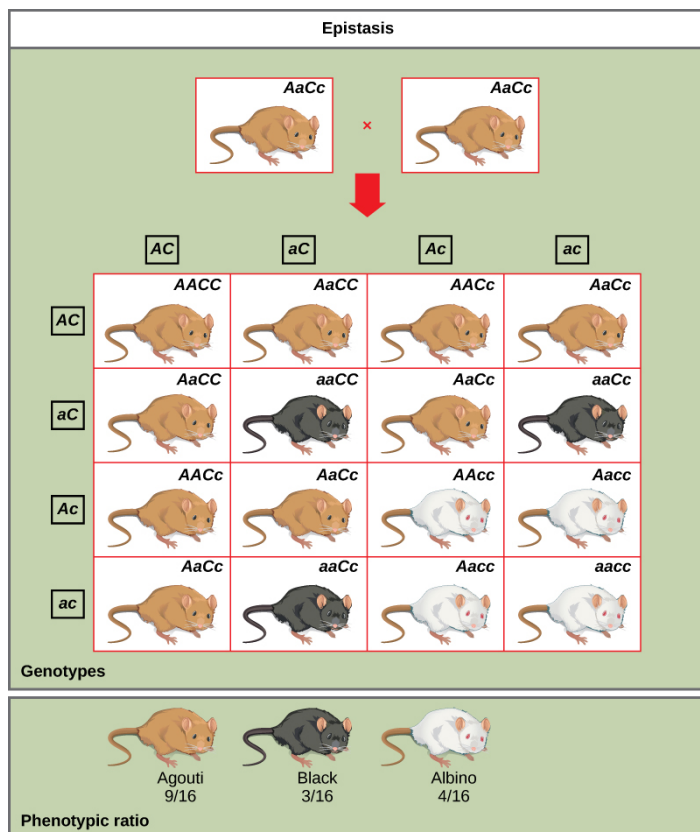
Eye color in humans is determined by multiple genes. Use the [Eye Color Calculator](#) to predict the eye color of children from parental eye color.

In some cases, several genes can contribute to aspects of a common phenotype without their gene products ever directly interacting. In the case of organ development, for instance, genes may be expressed sequentially, with each gene adding to the complexity and specificity of the organ. Genes may function in complementary or synergistic fashions, such that two or more genes need to be expressed simultaneously to affect a phenotype. Genes may also oppose each other, with one gene modifying the expression of another.

In **epistasis**, the interaction between genes is antagonistic, such that one gene masks or interferes with the expression of another. “Epistasis” is a

word composed of Greek roots that mean “standing upon.” The alleles that are being masked or silenced are said to be hypostatic to the epistatic alleles that are doing the masking. Often the biochemical basis of epistasis is a gene pathway in which the expression of one gene is dependent on the function of a gene that precedes or follows it in the pathway.

An example of epistasis is pigmentation in mice. The wild-type coat color, agouti (*AA*), is dominant to solid-colored fur (*aa*). However, a separate gene (*C*) is necessary for pigment production. A mouse with a recessive *c* allele at this locus is unable to produce pigment and is albino regardless of the allele present at locus *A* ([link](#)). Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce the same albino phenotype. A cross between heterozygotes for both genes (*AaCc* x *AaCc*) would generate offspring with a phenotypic ratio of 9 agouti:3 solid color:4 albino ([link](#)). In this case, the *C* gene is epistatic to the *A* gene.



In mice, the mottled agouti coat color (A) is dominant to a solid coloration, such as black or gray. A gene at a separate locus (C) is responsible for pigment production. The recessive *c* allele does not produce pigment, and a mouse with the homozygous recessive *cc* genotype is albino regardless of the allele present at the A locus. Thus, the C gene is epistatic to the A gene.

Epistasis can also occur when a dominant allele masks expression at a separate gene. Fruit color in summer squash is expressed in this way. Homozygous recessive expression of the W gene (*ww*) coupled with homozygous dominant or heterozygous expression of the Y gene (*YY* or *Yy*) generates yellow fruit, and the *wwyy* genotype produces green fruit. However, if a dominant copy of the W gene is present in the homozygous or heterozygous form, the summer squash will produce white fruit regardless of the Y alleles. A cross between white heterozygotes for both genes (*WwYy* × *WwYy*) would produce offspring with a phenotypic ratio of 12 white:3 yellow:1 green.

Finally, epistasis can be reciprocal such that either gene, when present in the dominant (or recessive) form, expresses the same phenotype. In the shepherd's purse plant (*Capsella bursa-pastoris*), the characteristic of seed shape is controlled by two genes in a dominant epistatic relationship. When the genes A and B are both homozygous recessive (*aabb*), the seeds are ovoid. If the dominant allele for either of these genes is present, the result is triangular seeds. That is, every possible genotype other than *aabb* results in triangular seeds, and a cross between heterozygotes for both genes (*AaBb* × *AaBb*) would yield offspring with a phenotypic ratio of 15 triangular:1 ovoid.

As you work through genetics problems, keep in mind that any single characteristic that results in a phenotypic ratio that totals 16 is typical of a

two-gene interaction. Recall the phenotypic inheritance pattern for Mendel's dihybrid cross, which considered two non-interacting genes—9:3:3:1. Similarly, we would expect interacting gene pairs to also exhibit ratios expressed as 16 parts. Note that we are assuming the interacting genes are not linked; they are still assorting independently into gametes.

Note:

Link to Learning



For an excellent review of Mendel's experiments and to perform your own crosses and identify patterns of inheritance, visit the [Mendel's Peas](#) web lab.

Section Summary

Mendel postulated that genes (characteristics) are inherited as pairs of alleles (traits) that behave in a dominant and recessive pattern. Alleles segregate into gametes such that each gamete is equally likely to receive either one of the two alleles present in a diploid individual. In addition, genes are assorted into gametes independently of one another. That is, alleles are generally not more likely to segregate into a gamete with a particular allele of another gene. A dihybrid cross demonstrates independent assortment when the genes in question are on different chromosomes or distant from each other on the same chromosome. For crosses involving more than two genes, use the forked line or probability methods to predict offspring genotypes and phenotypes rather than a Punnett square.

Although chromosomes sort independently into gametes during meiosis, Mendel's law of independent assortment refers to genes, not chromosomes, and a single chromosome may carry more than 1,000 genes. When genes are located in close proximity on the same chromosome, their alleles tend to be inherited together. This results in offspring ratios that violate Mendel's law of independent assortment. However, recombination serves to exchange genetic material on homologous chromosomes such that maternal and paternal alleles may be recombined on the same chromosome. This is why alleles on a given chromosome are not always inherited together. Recombination is a random event occurring anywhere on a chromosome. Therefore, genes that are far apart on the same chromosome are likely to still assort independently because of recombination events that occurred in the intervening chromosomal space.

Whether or not they are sorting independently, genes may interact at the level of gene products such that the expression of an allele for one gene masks or modifies the expression of an allele for a different gene. This is called epistasis.

Art Connections

Exercise:

Problem:

[\[link\]](#) In pea plants, purple flowers (P) are dominant to white flowers (p) and yellow peas (Y) are dominant to green peas (y). What are the possible genotypes and phenotypes for a cross between PpYY and ppYy pea plants? How many squares do you need to do a Punnett square analysis of this cross?

Solution:

[\[link\]](#) The possible genotypes are PpYY, PpYy, ppYY, and ppYy. The former two genotypes would result in plants with purple flowers and yellow peas, while the latter two genotypes would result in plants with white flowers with yellow peas, for a 1:1 ratio of each phenotype. You

only need a 2×2 Punnett square (four squares total) to do this analysis because two of the alleles are homozygous.

Multiple Choice

Exercise:

Problem:

Assuming no gene linkage, in a dihybrid cross of $AABB \times aabb$ with $AaBb$ F_1 heterozygotes, what is the ratio of the F_1 gametes (AB , aB , Ab , ab) that will give rise to the F_2 offspring?

- a. 1:1:1:1
- b. 1:3:3:1
- c. 1:2:2:1
- d. 4:3:2:1

Solution:

A

Exercise:

Problem:

The forked line and probability methods make use of what probability rule?

- a. test cross
- b. product rule
- c. monohybrid rule
- d. sum rule

Solution:

B

Exercise:**Problem:**

How many different offspring genotypes are expected in a trihybrid cross between parents heterozygous for all three traits when the traits behave in a dominant and recessive pattern? How many phenotypes?

- a. 64 genotypes; 16 phenotypes
- b. 16 genotypes; 64 phenotypes
- c. 8 genotypes; 27 phenotypes
- d. 27 genotypes; 8 phenotypes

Solution:

D

Free Response**Exercise:****Problem:**

Use the probability method to calculate the genotypes and genotypic proportions of a cross between *AABBCc* and *Aabbcc* parents.

Solution:

Considering each gene separately, the cross at *A* will produce offspring of which half are *AA* and half are *Aa*; *B* will produce all *Bb*; *C* will produce half *Cc* and half *cc*. Proportions then are $(1/2) \times (1) \times (1/2)$, or $1/4$ *AABbCc*; continuing for the other possibilities yields $1/4$ *AABbcc*, $1/4$ *AaBbCc*, and $1/4$ *AaBbcc*. The proportions therefore are 1:1:1:1.

Exercise:

Problem:

Explain epistasis in terms of its Greek-language roots “standing upon.”

Solution:

Epistasis describes an antagonistic interaction between genes wherein one gene masks or interferes with the expression of another. The gene that is interfering is referred to as epistatic, as if it is “standing upon” the other (hypostatic) gene to block its expression.

Exercise:**Problem:**

In Section 12.3, “Laws of Inheritance,” an example of epistasis was given for the summer squash. Cross white $WwYy$ heterozygotes to prove the phenotypic ratio of 12 white:3 yellow:1 green that was given in the text.

Solution:

The cross can be represented as a 4×4 Punnett square, with the following gametes for each parent: WY , Wy , wY , and wy . For all 12 of the offspring that express a dominant W gene, the offspring will be white. The three offspring that are homozygous recessive for w but express a dominant Y gene will be yellow. The remaining $wwyy$ offspring will be green.

Glossary**dihybrid**

result of a cross between two true-breeding parents that express different traits for two characteristics

epistasis

antagonistic interaction between genes such that one gene masks or interferes with the expression of another

law of dominance

in a heterozygote, one trait will conceal the presence of another trait for the same characteristic

law of independent assortment

genes do not influence each other with regard to sorting of alleles into gametes; every possible combination of alleles is equally likely to occur

law of segregation

paired unit factors (i.e., genes) segregate equally into gametes such that offspring have an equal likelihood of inheriting any combination of factors

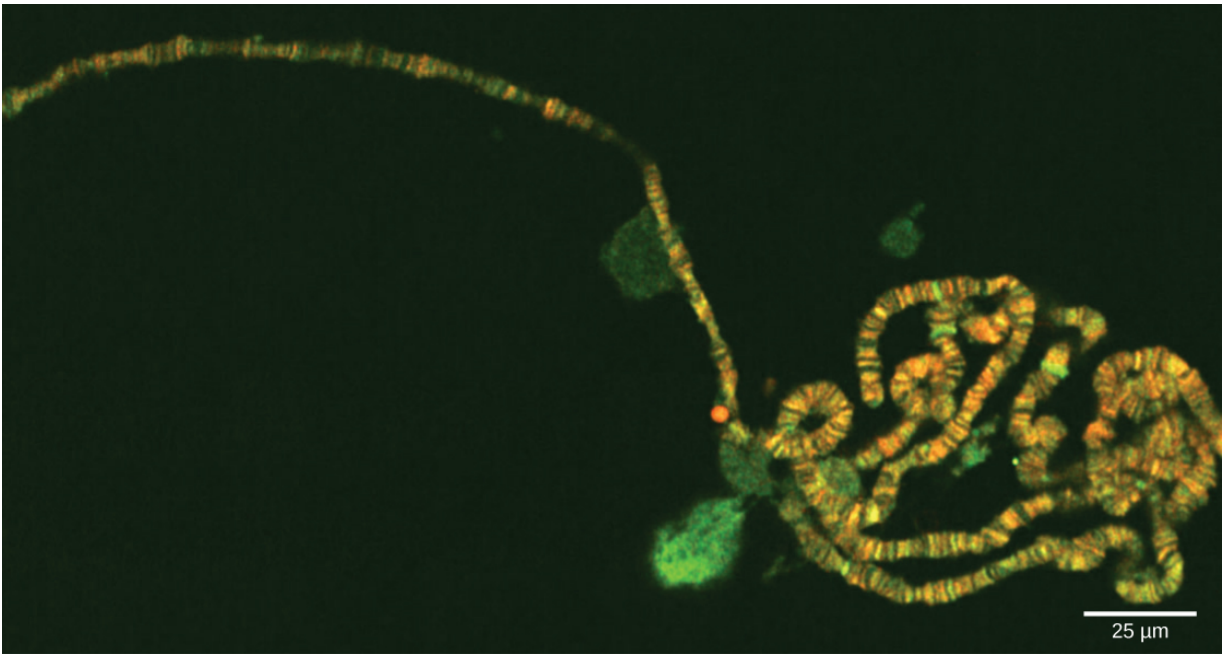
linkage

phenomenon in which alleles that are located in close proximity to each other on the same chromosome are more likely to be inherited together

Introduction

class="introduction"

Chromosomes are threadlike nuclear structures consisting of DNA and proteins that serve as the repositories for genetic information. The chromosomes depicted here were isolated from a fruit fly's salivary gland, stained with dye, and visualized under a microscope. Akin to miniature bar codes, chromosomes absorb different dyes to produce characteristic banding patterns, which allows for their routine identification. (credit: modification of work by "LPLT"/Wikimedia Commons; scale-bar data from Matt Russell)



The gene is the physical unit of inheritance, and genes are arranged in a linear order on chromosomes. The behaviors and interactions of chromosomes during meiosis explain, at a cellular level, the patterns of inheritance that we observe in populations. Genetic disorders involving alterations in chromosome number or structure may have dramatic effects and can prevent a fertilized egg from developing altogether.

Chromosomal Theory and Genetic Linkage

By the end of this section, you will be able to:

- Discuss Sutton's Chromosomal Theory of Inheritance
- Describe genetic linkage
- Explain the process of homologous recombination, or crossing over
- Describe how chromosome maps are created
- Calculate the distances between three genes on a chromosome using a three-point test cross

Long before chromosomes were visualized under a microscope, the father of modern genetics, Gregor Mendel, began studying heredity in 1843. With the improvement of microscopic techniques during the late 1800s, cell biologists could stain and visualize subcellular structures with dyes and observe their actions during cell division and meiosis. With each mitotic division, chromosomes replicated, condensed from an amorphous (no constant shape) nuclear mass into distinct X-shaped bodies (pairs of identical sister chromatids), and migrated to separate cellular poles.

Chromosomal Theory of Inheritance

The speculation that chromosomes might be the key to understanding heredity led several scientists to examine Mendel's publications and re-evaluate his model in terms of the behavior of chromosomes during mitosis and meiosis. In 1902, Theodor Boveri observed that proper embryonic development of sea urchins does not occur unless chromosomes are present. That same year, Walter Sutton observed the separation of chromosomes into daughter cells during meiosis ([\[link\]](#)). Together, these observations led to the development of the **Chromosomal Theory of Inheritance**, which identified chromosomes as the genetic material responsible for Mendelian inheritance.



(a) Walter Sutton and (b) Theodor Boveri are credited with developing the Chromosomal Theory of Inheritance, which states that chromosomes carry the unit of heredity (genes).

The Chromosomal Theory of Inheritance was consistent with Mendel's laws and was supported by the following observations:

- During meiosis, homologous chromosome pairs migrate as discrete structures that are independent of other chromosome pairs.
- The sorting of chromosomes from each homologous pair into pre-gametes appears to be random.
- Each parent synthesizes gametes that contain only half of their chromosomal complement.
- Even though male and female gametes (sperm and egg) differ in size and morphology, they have the same number of chromosomes, suggesting equal genetic contributions from each parent.
- The gametic chromosomes combine during fertilization to produce offspring with the same chromosome number as their parents.

Despite compelling correlations between the behavior of chromosomes during meiosis and Mendel's abstract laws, the Chromosomal Theory of

Inheritance was proposed long before there was any direct evidence that traits were carried on chromosomes. Critics pointed out that individuals had far more independently segregating traits than they had chromosomes. It was only after several years of carrying out crosses with the fruit fly, *Drosophila melanogaster*, that Thomas Hunt Morgan provided experimental evidence to support the Chromosomal Theory of Inheritance.

Genetic Linkage and Distances

Mendel's work suggested that traits are inherited independently of each other. Morgan identified a 1:1 correspondence between a segregating trait and the X chromosome, suggesting that the random segregation of chromosomes was the physical basis of Mendel's model. This also demonstrated that linked genes disrupt Mendel's predicted outcomes. The fact that each chromosome can carry many linked genes explains how individuals can have many more traits than they have chromosomes. However, observations by researchers in Morgan's laboratory suggested that alleles positioned on the same chromosome were not always inherited together. During meiosis, linked genes somehow became unlinked.

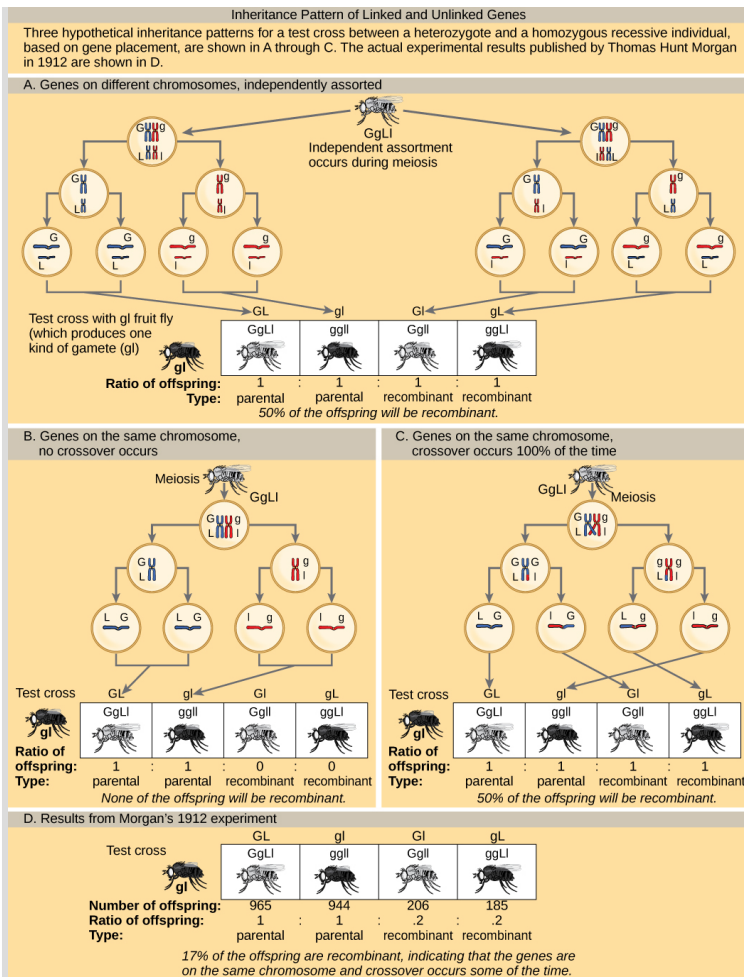
Homologous Recombination

In 1909, Frans Janssen observed chiasmata—the point at which chromatids are in contact with each other and may exchange segments—prior to the first division of meiosis. He suggested that alleles become unlinked and chromosomes physically exchange segments. As chromosomes condensed and paired with their homologs, they appeared to interact at distinct points. Janssen suggested that these points corresponded to regions in which chromosome segments were exchanged. It is now known that the pairing and interaction between homologous chromosomes, known as synapsis, does more than simply organize the homologs for migration to separate daughter cells. When synapsed, homologous chromosomes undergo reciprocal physical exchanges at their arms in a process called **homologous recombination**, or more simply, “crossing over.”

To better understand the type of experimental results that researchers were obtaining at this time, consider a heterozygous individual that inherited dominant maternal alleles for two genes on the same chromosome (such as *AB*) and two recessive paternal alleles for those same genes (such as *ab*). If the genes are linked, one would expect this individual to produce gametes that are either *AB* or *ab* with a 1:1 ratio. If the genes are unlinked, the individual should produce *AB*, *Ab*, *aB*, and *ab* gametes with equal frequencies, according to the Mendelian concept of independent assortment. Because they correspond to new allele combinations, the genotypes *Ab* and *aB* are **nonparental types** that result from homologous recombination during meiosis. **Parental types** are progeny that exhibit the same allelic combination as their parents. Morgan and his colleagues, however, found that when such heterozygous individuals were test crossed to a homozygous recessive parent ($AaBb \times aabb$), both parental and nonparental cases occurred. For example, 950 offspring might be recovered that were either *AaBb* or *aabb*, but 50 offspring would also be obtained that were either *Aabb* or *aaBb*. These results suggested that linkage occurred most often, but a significant minority of offspring were the products of recombination.

Note:

Art Connection



Inheritance patterns of unlinked and linked genes are shown. In (a), two genes are located on different chromosomes so independent assortment occurs during meiosis. The offspring have an equal chance of being the parental type (inheriting the same combination of traits as the parents) or a nonparental type (inheriting a different combination of traits than the parents). In (b), two genes are very close together on the same chromosome so that no crossing over occurs between them. The genes are therefore always inherited together and all of the offspring are the

parental type. In (c), two genes are far apart on the chromosome such that crossing over occurs during every meiotic event. The recombination frequency will be the same as if the genes were on separate chromosomes. (d) The actual recombination frequency of fruit fly wing length and body color that Thomas Morgan observed in 1912 was 17 percent. A crossover frequency between 0 percent and 50 percent indicates that the genes are on the same chromosome and crossover occurs some of the time.

In a test cross for two characteristics such as the one shown here, can the predicted frequency of recombinant offspring be 60 percent? Why or why not?

Genetic Maps

Janssen did not have the technology to demonstrate crossing over so it remained an abstract idea that was not widely accepted. Scientists thought chiasmata were a variation on synapsis and could not understand how chromosomes could break and rejoin. Yet, the data were clear that linkage did not always occur. Ultimately, it took a young undergraduate student and an “all-nighter” to mathematically elucidate the problem of linkage and recombination.

In 1913, Alfred Sturtevant, a student in Morgan’s laboratory, gathered results from researchers in the laboratory, and took them home one night to mull them over. By the next morning, he had created the first “chromosome map,” a linear representation of gene order and relative distance on a chromosome ([link](#)).

Note:

Art Connection

Genetic Map Based on Recombination Frequencies in <i>Drosophila</i>			
MUTANT			WILD TYPE
Short aristae	0		Long aristae
Black body	48.5		Gray body
Cinnabar eyes	57.5		Red eyes
Vestigial wings	65.5		Normal wings
Brown eyes	104.5		Red eyes
Values in centimorgan (cM) map units; recombination frequency of 0.01 = 1 cM			

This genetic map orders *Drosophila* genes on the basis of recombination frequency.

Which of the following statements is true?

- a. Recombination of the body color and red/cinnabar eye alleles will occur more frequently than recombination of the alleles for wing length and aristae length.
- b. Recombination of the body color and aristae length alleles will occur more frequently than recombination of red/brown eye alleles and the aristae length alleles.
- c. Recombination of the gray/black body color and long/short aristae alleles will not occur.
- d. Recombination of the red/brown eye and long/short aristae alleles will occur more frequently than recombination of the alleles for wing length and body color.

As shown in [\[link\]](#), by using recombination frequency to predict genetic distance, the relative order of genes on chromosome 2 could be inferred. The values shown represent map distances in centimorgans (cM), which correspond to recombination frequencies (in percent). Therefore, the genes for body color and wing size were $65.5 - 48.5 = 17$ cM apart, indicating that the maternal and paternal alleles for these genes recombine in 17 percent of offspring, on average.

To construct a chromosome map, Sturtevant assumed that genes were ordered serially on threadlike chromosomes. He also assumed that the incidence of recombination between two homologous chromosomes could occur with equal likelihood anywhere along the length of the chromosome. Operating under these assumptions, Sturtevant postulated that alleles that were far apart on a chromosome were more likely to dissociate during meiosis simply because there was a larger region over which recombination could occur. Conversely, alleles that were close to each other on the chromosome were likely to be inherited together. The average number of crossovers between two alleles—that is, their **recombination frequency**—correlated with their genetic distance from each other, relative to the locations of other genes on that chromosome. Considering the example cross between *AaBb* and *aabb* above, the frequency of recombination could be calculated as $50/1000 = 0.05$. That is, the likelihood of a crossover between genes *A/a* and *B/b* was 0.05, or 5 percent. Such a result would indicate that the genes were definitively linked, but that they were far enough apart for crossovers to occasionally occur. Sturtevant divided his genetic map into map units, or **centimorgans (cM)**, in which a recombination frequency of 0.01 corresponds to 1 cM.

By representing alleles in a linear map, Sturtevant suggested that genes can range from being perfectly linked (recombination frequency = 0) to being perfectly unlinked (recombination frequency = 0.5) when genes are on different chromosomes or genes are separated very far apart on the same chromosome. Perfectly unlinked genes correspond to the frequencies predicted by Mendel to assort independently in a dihybrid cross. A recombination frequency of 0.5 indicates that 50 percent of offspring are recombinants and the other 50 percent are parental types. That is, every type of allele combination is represented with equal frequency. This

representation allowed Sturtevant to additively calculate distances between several genes on the same chromosome. However, as the genetic distances approached 0.50, his predictions became less accurate because it was not clear whether the genes were very far apart on the same chromosome or on different chromosomes.

In 1931, Barbara McClintock and Harriet Creighton demonstrated the crossover of homologous chromosomes in corn plants. Weeks later, homologous recombination in *Drosophila* was demonstrated microscopically by Curt Stern. Stern observed several X-linked phenotypes that were associated with a structurally unusual and dissimilar X chromosome pair in which one X was missing a small terminal segment, and the other X was fused to a piece of the Y chromosome. By crossing flies, observing their offspring, and then visualizing the offspring's chromosomes, Stern demonstrated that every time the offspring allele combination deviated from either of the parental combinations, there was a corresponding exchange of an X chromosome segment. Using mutant flies with structurally distinct X chromosomes was the key to observing the products of recombination because DNA sequencing and other molecular tools were not yet available. It is now known that homologous chromosomes regularly exchange segments in meiosis by reciprocally breaking and rejoining their DNA at precise locations.

Note:

Link to Learning



Review Sturtevant's process to create a genetic map on the basis of recombination frequencies [here](#).

Mendel's Mapped Traits

Homologous recombination is a common genetic process, yet Mendel never observed it. Had he investigated both linked and unlinked genes, it would have been much more difficult for him to create a unified model of his data on the basis of probabilistic calculations. Researchers who have since mapped the seven traits investigated by Mendel onto the seven chromosomes of the pea plant genome have confirmed that all of the genes he examined are either on separate chromosomes or are sufficiently far apart as to be statistically unlinked. Some have suggested that Mendel was enormously lucky to select only unlinked genes, whereas others question whether Mendel discarded any data suggesting linkage. In any case, Mendel consistently observed independent assortment because he examined genes that were effectively unlinked.

Section Summary

The Chromosomal Theory of inheritance, proposed by Sutton and Boveri, states that chromosomes are the vehicles of genetic heredity. Neither Mendelian genetics nor gene linkage is perfectly accurate; instead, chromosome behavior involves segregation, independent assortment, and occasionally, linkage. Sturtevant devised a method to assess recombination frequency and infer the relative positions and distances of linked genes on a chromosome on the basis of the average number of crossovers in the intervening region between the genes. Sturtevant correctly presumed that genes are arranged in serial order on chromosomes and that recombination between homologs can occur anywhere on a chromosome with equal likelihood. Whereas linkage causes alleles on the same chromosome to be inherited together, homologous recombination biases alleles toward an inheritance pattern of independent assortment.

Art Connections

Exercise:

Problem:

[\[link\]](#) In a test cross for two characteristics such as the one shown here, can the predicted frequency of recombinant offspring be 60 percent? Why or why not?

Solution:

[\[link\]](#) No. The predicted frequency of recombinant offspring ranges from 0% (for linked traits) to 50% (for unlinked traits).

Exercise:

Problem: [\[link\]](#) Which of the following statements is true?

- a. Recombination of the body color and red/cinnabar eye alleles will occur more frequently than recombination of the alleles for wing length and aristae length.
 - b. Recombination of the body color and aristae length alleles will occur more frequently than recombination of red/brown eye alleles and the aristae length alleles.
 - c. Recombination of the gray/black body color and long/short aristae alleles will not occur.
 - d. Recombination of the red/brown eye and long/short aristae alleles will occur more frequently than recombination of the alleles for wing length and body color.
-

Solution:

[\[link\]](#) D

Review Questions**Exercise:**

Problem:

X-linked recessive traits in humans (or in *Drosophila*) are observed _____.

- a. in more males than females
- b. in more females than males
- c. in males and females equally
- d. in different distributions depending on the trait

Solution:

A

Exercise:**Problem:**

The first suggestion that chromosomes may physically exchange segments came from the microscopic identification of _____.

- a. synapsis
- b. sister chromatids
- c. chiasmata
- d. alleles

Solution:

C

Exercise:**Problem:**

Which recombination frequency corresponds to independent assortment and the absence of linkage?

- a. 0

- b. 0.25
- c. 0.50
- d. 0.75

Solution:

C

Exercise:

Problem:

Which recombination frequency corresponds to perfect linkage and violates the law of independent assortment?

- a. 0
- b. 0.25
- c. 0.50
- d. 0.75

Solution:

A

Free Response

Exercise:

Problem:

Explain how the Chromosomal Theory of Inheritance helped to advance our understanding of genetics.

Solution:

The Chromosomal Theory of Inheritance proposed that genes reside on chromosomes. The understanding that chromosomes are linear arrays

of genes explained linkage, and crossing over explained recombination.

Glossary

centimorgan (cM)

(also, map unit) relative distance that corresponds to a recombination frequency of 0.01

Chromosomal Theory of Inheritance

theory proposing that chromosomes are the vehicles of genes and that their behavior during meiosis is the physical basis of the inheritance patterns that Mendel observed

homologous recombination

process by which homologous chromosomes undergo reciprocal physical exchanges at their arms, also known as crossing over

nonparental (recombinant) type

progeny resulting from homologous recombination that exhibits a different allele combination compared with its parents

parental types

progeny that exhibits the same allelic combination as its parents

recombination frequency

average number of crossovers between two alleles; observed as the number of nonparental types in a population of progeny

Genetic Linkages and Distances

Chromosomal Basis of Inherited Disorders

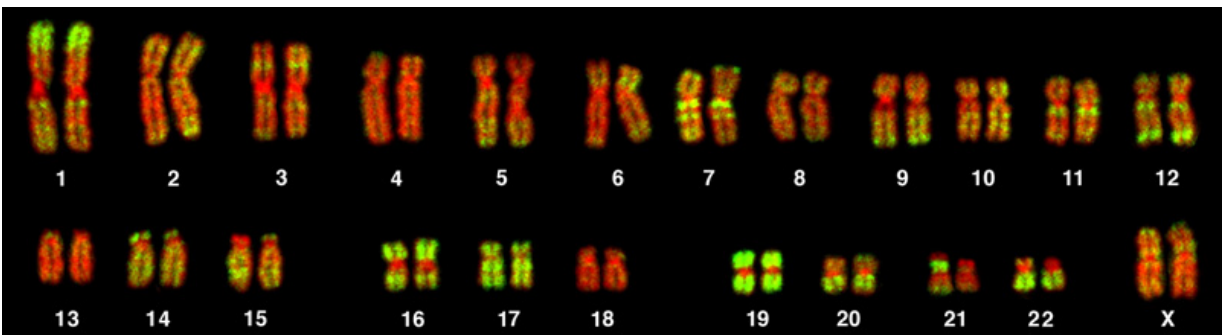
By the end of this section, you will be able to:

- Describe how a karyogram is created
- Explain how nondisjunction leads to disorders in chromosome number
- Compare disorders caused by aneuploidy
- Describe how errors in chromosome structure occur through inversions and translocations

Inherited disorders can arise when chromosomes behave abnormally during meiosis. Chromosome disorders can be divided into two categories: abnormalities in chromosome number and chromosomal structural rearrangements. Because even small segments of chromosomes can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Identification of Chromosomes

The isolation and microscopic observation of chromosomes forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans. A **karyotype** is the number and appearance of chromosomes, and includes their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or **karyogram**, also known as an ideogram ([link](#)).



This karyotype is of a female human. Notice that homologous chromosomes are the same size, and have the same centromere positions and banding patterns. A human male would have an XY chromosome pair instead of the XX pair shown. (credit: Andreas Blozer et al)

In a given species, chromosomes can be identified by their number, size, centromere position, and banding pattern. In a human karyotype, **autosomes** or “body chromosomes” (all of the non–sex chromosomes) are generally organized in approximate order of size from largest (chromosome 1) to smallest (chromosome 22). The X and Y chromosomes are not autosomes. However, chromosome 21 is actually shorter than chromosome 22. This was discovered after the naming of Down syndrome as trisomy 21, reflecting how this disease results from possessing one extra chromosome 21 (three total). Not wanting to change the name of this important disease, chromosome 21 retained its numbering, despite describing the shortest set of chromosomes. The chromosome “arms” projecting from either end of the centromere may be designated as short or long, depending on their relative lengths. The short arm is abbreviated *p* (for “petite”), whereas the long arm is abbreviated *q* (because it follows “p” alphabetically). Each arm is further subdivided and denoted by a number. Using this naming system, locations on chromosomes can be described consistently in the scientific literature.

Note:**Career Connection****Geneticists Use Karyograms to Identify Chromosomal Aberrations**

Although Mendel is referred to as the “father of modern genetics,” he performed his experiments with none of the tools that the geneticists of today routinely employ. One such powerful cytological technique is karyotyping, a method in which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual’s karyotype, a person’s cells (like white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated

cells are stimulated to begin actively dividing. A chemical called colchicine is then applied to cells to arrest condensed chromosomes in metaphase. Cells are then made to swell using a hypotonic solution so the chromosomes spread apart. Finally, the sample is preserved in a fixative and applied to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, the chromosomes are viewed using bright-field microscopy. A common stain choice is the Giemsa stain.

Giemsa staining results in approximately 400–800 bands (of tightly coiled DNA and condensed proteins) arranged along all of the 23 chromosome pairs; an experienced geneticist can identify each band. In addition to the banding patterns, chromosomes are further identified on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous pairs of chromosomes are aligned in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern ([link](#)).

At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are Down Syndrome, which is identified by a third copy of chromosome 21, and Turner Syndrome, which is characterized by the presence of only one X chromosome in women instead of the normal two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen Syndrome—which involves distinctive facial features as well as heart and bleeding defects—is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint **translocations**, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in certain cancers, including chronic myelogenous leukemia.

During Mendel's lifetime, inheritance was an abstract concept that could only be inferred by performing crosses and observing the traits expressed by offspring. By observing a karyogram, today's geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring, even before birth.

Disorders in Chromosome Number

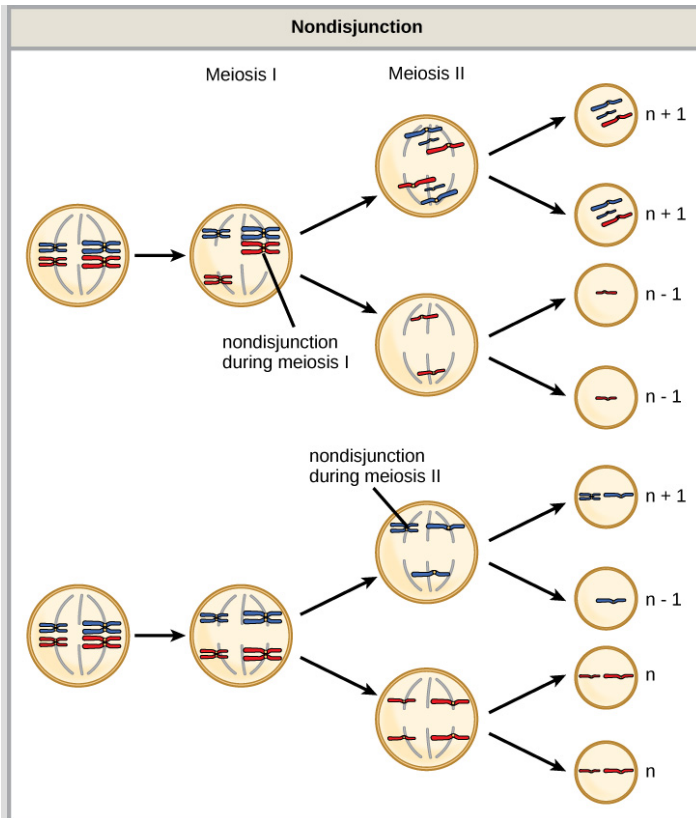
Of all of the chromosomal disorders, abnormalities in chromosome number are the most obviously identifiable from a karyogram. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis.

Misaligned or incomplete synapsis, or a dysfunction of the spindle apparatus that facilitates chromosome migration, can cause nondisjunction. The risk of nondisjunction occurring increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II, with differing results ([link](#)). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that particular chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.

Note:

Art Connection



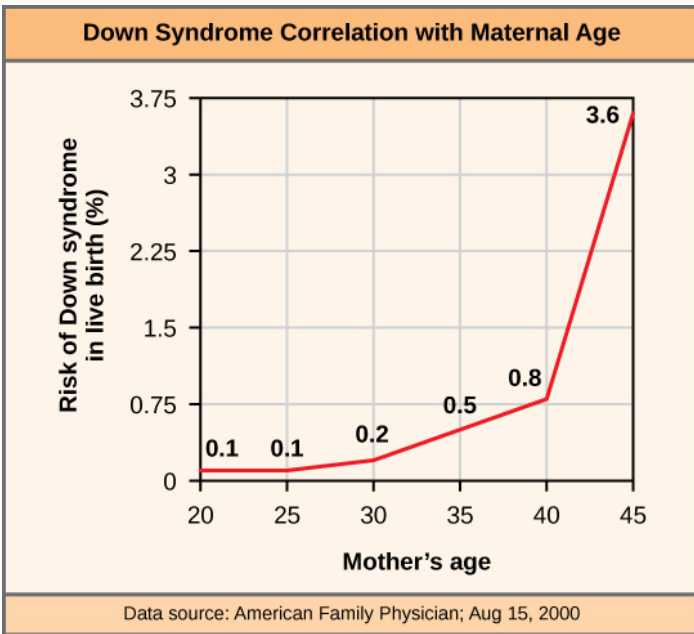
Nondisjunction occurs when homologous chromosomes or sister chromatids fail to separate during meiosis, resulting in an abnormal chromosome number. Nondisjunction may occur during meiosis I or meiosis II.

Which of the following statements about nondisjunction is true?

- Nondisjunction only results in gametes with $n+1$ or $n-1$ chromosomes.
- Nondisjunction occurring during meiosis II results in 50 percent normal gametes.
- Nondisjunction during meiosis I results in 50 percent normal gametes.
- Nondisjunction always results in four different kinds of gametes.

Aneuploidy

An individual with the appropriate number of chromosomes for their species is called **euploid**; in humans, euploidy corresponds to 22 pairs of autosomes and one pair of sex chromosomes. An individual with an error in chromosome number is described as **aneuploid**, a term that includes **monosomy** (loss of one chromosome) or **trisomy** (gain of an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they lack essential genes. This underscores the importance of “gene dosage” in humans. Most autosomal trisomies also fail to develop to birth; however, duplications of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Individuals with an extra chromosome may synthesize an abundance of the gene products encoded by that chromosome. This extra dose (150 percent) of specific genes can lead to a number of functional challenges and often precludes development. The most common trisomy among viable births is that of chromosome 21, which corresponds to Down Syndrome. Individuals with this inherited disorder are characterized by short stature and stunted digits, facial distinctions that include a broad skull and large tongue, and significant developmental delays. The incidence of Down syndrome is correlated with maternal age; older women are more likely to become pregnant with fetuses carrying the trisomy 21 genotype ([\[link\]](#)).



The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

Note:

Link to Learning



Visualize the addition of a chromosome that leads to Down syndrome in this [video simulation](#).

Polyploidy

An individual with more than the correct number of chromosome sets (two for diploid species) is called **polyploid**. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the flatworms, crustaceans, amphibians, fish, and lizards. Polyploid animals are sterile because meiosis cannot proceed normally and instead produces mostly aneuploid daughter cells that cannot yield viable zygotes. Rarely, polyploid animals can reproduce asexually by haplodiploidy, in which an unfertilized egg divides mitotically to produce offspring. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species ([link](#)).



As with many polyploid plants, this triploid orange daylily (*Hemerocallis fulva*) is particularly large and robust, and grows flowers with triple the number of petals of its diploid counterparts. (credit: Steve Karg)

Sex Chromosome Nondisjunction in Humans

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally, despite carrying different numbers of the X chromosome. Rather than a gain or loss of autosomes, variations in the number of sex chromosomes are associated with relatively mild effects. In part, this occurs because of a molecular process called **X inactivation**. Early in development, when female mammalian embryos consist of just a few thousand cells (relative to trillions in the newborn), one X chromosome in each cell inactivates by tightly condensing into a quiescent (dormant) structure called a Barr body. The chance that an X chromosome (maternally or paternally derived) is inactivated in each cell is random, but once the inactivation occurs, all cells derived from that one will have the same inactive X chromosome or Barr body. By this process, females compensate for their double genetic dose of X chromosome. In so-called “tortoiseshell” cats, embryonic X inactivation is observed as color variegation ([\[link\]](#)). Females that are heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region.



In cats, the gene for

coat color is located on the X chromosome.

In the embryonic development of female cats, one of the two X chromosomes is randomly inactivated in each cell, resulting in a tortoiseshell pattern if the cat has two different alleles for coat color. Male cats, having only one X chromosome, never exhibit a tortoiseshell coat color. (credit: Michael Bodega)

An individual carrying an abnormal number of X chromosomes will inactivate all but one X chromosome in each of her cells. However, even inactivated X chromosomes continue to express a few genes, and X chromosomes must reactivate for the proper maturation of female ovaries. As a result, X-chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop in utero.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes, called triplo-X, are phenotypically female but express developmental delays and reduced fertility. The XXY genotype, corresponding to one type of Klinefelter syndrome, corresponds to phenotypically male individuals with small testes, enlarged breasts, and reduced body hair. More complex types of Klinefelter syndrome exist in which the individual has as many as five X chromosomes. In all types, every X chromosome except one undergoes inactivation to compensate for the excess genetic dosage. This can be seen as several Barr bodies in each

cell nucleus. Turner syndrome, characterized as an XO genotype (i.e., only a single sex chromosome), corresponds to a phenotypically female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

Duplications and Deletions

In addition to the loss or gain of an entire chromosome, a chromosomal segment may be duplicated or lost. Duplications and deletions often produce offspring that survive but exhibit physical and mental abnormalities. Duplicated chromosomal segments may fuse to existing chromosomes or may be free in the nucleus. Cri-du-chat (from the French for “cry of the cat”) is a syndrome associated with nervous system abnormalities and identifiable physical features that result from a deletion of most of 5p (the small arm of chromosome 5) ([link](#)). Infants with this genotype emit a characteristic high-pitched cry on which the disorder’s name is based.



This individual with cri-du-chat syndrome is shown at two, four, nine, and 12 years of age. (credit: Paola Cerruti Mainardi)

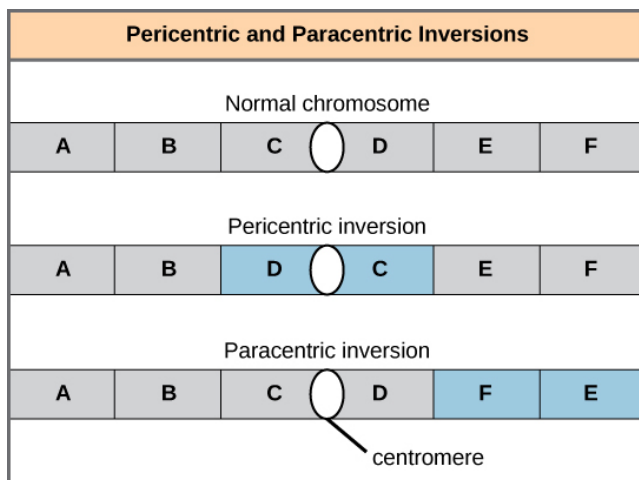
Chromosomal Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, but chromosome inversions and translocations are the most common. Both are identified during meiosis by the adaptive pairing of rearranged chromosomes with their former homologs to maintain appropriate gene alignment. If the genes carried on two homologs are not oriented correctly, a recombination event could result in the loss of genes from one chromosome and the gain of genes on the other. This would produce aneuploid gametes.

Chromosome Inversions

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome. Inversions may occur in nature as a result of mechanical shear, or from the action of transposable elements (special DNA sequences capable of facilitating the rearrangement of chromosome segments with the help of enzymes that cut and paste DNA sequences). Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have more mild effects than aneuploid errors. However, altered gene orientation can result in functional changes because regulators of gene expression could be moved out of position with respect to their targets, causing aberrant levels of gene products.

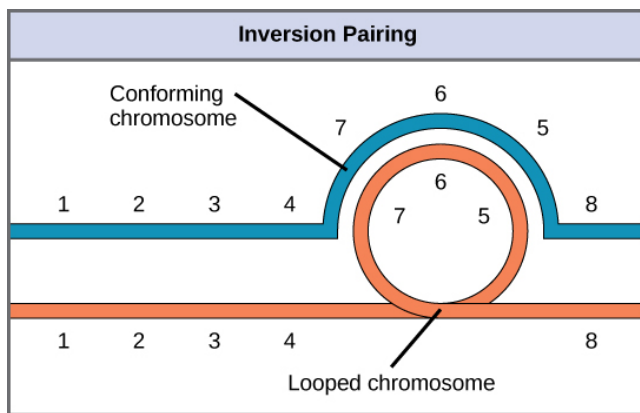
An inversion can be **pericentric** and include the centromere, or **paracentric** and occur outside of the centromere ([\[link\]](#)). A pericentric inversion that is asymmetric about the centromere can change the relative lengths of the chromosome arms, making these inversions easily identifiable.



Pericentric inversions include the centromere, and paracentric inversions do not. A pericentric inversion can change the relative

lengths of the chromosome arms;
a paracentric inversion cannot.

When one homologous chromosome undergoes an inversion but the other does not, the individual is described as an inversion heterozygote. To maintain point-for-point synapsis during meiosis, one homolog must form a loop, and the other homolog must mold around it. Although this topology can ensure that the genes are correctly aligned, it also forces the homologs to stretch and can be associated with regions of imprecise synapsis ([\[link\]](#)).



When one chromosome undergoes an inversion but the other does not, one chromosome must form an inverted loop to retain point-for-point interaction during synapsis. This inversion pairing is essential to maintaining gene alignment during meiosis and to allow for recombination.

Note:

Evolution Connection

The Chromosome 18 Inversion

Not all structural rearrangements of chromosomes produce nonviable, impaired, or infertile individuals. In rare instances, such a change can result in the evolution of a new species. In fact, a pericentric inversion in chromosome 18 appears to have contributed to the evolution of humans. This inversion is not present in our closest genetic relatives, the chimpanzees. Humans and chimpanzees differ cytogenetically by pericentric inversions on several chromosomes and by the fusion of two separate chromosomes in chimpanzees that correspond to chromosome two in humans.

The pericentric chromosome 18 inversion is believed to have occurred in early humans following their divergence from a common ancestor with chimpanzees approximately five million years ago. Researchers characterizing this inversion have suggested that approximately 19,000 nucleotide bases were duplicated on 18p, and the duplicated region inverted and reinserted on chromosome 18 of an ancestral human.

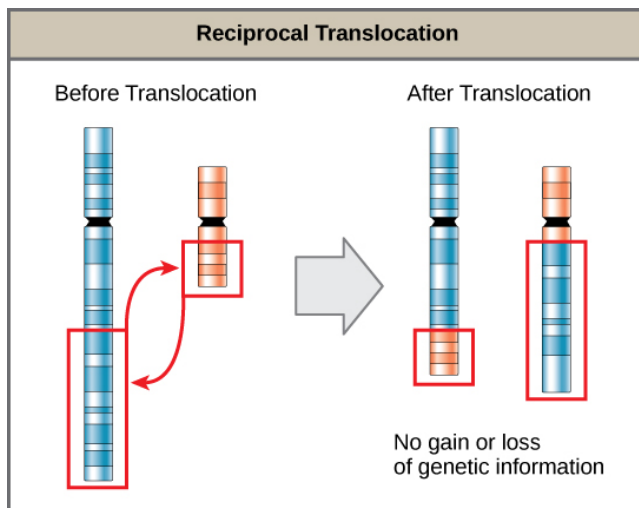
A comparison of human and chimpanzee genes in the region of this inversion indicates that two genes—*ROCK1* and *USP14*—that are adjacent on chimpanzee chromosome 17 (which corresponds to human chromosome 18) are more distantly positioned on human chromosome 18. This suggests that one of the inversion breakpoints occurred between these two genes.

Interestingly, humans and chimpanzees express *USP14* at distinct levels in specific cell types, including cortical cells and fibroblasts. Perhaps the chromosome 18 inversion in an ancestral human repositioned specific genes and reset their expression levels in a useful way. Because both *ROCK1* and *USP14* encode cellular enzymes, a change in their expression could alter cellular function. It is not known how this inversion contributed to hominid evolution, but it appears to be a significant factor in the divergence of humans from other primates. [\[footnote\]](#)

Violaine Goidts et al., “Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates,” *Human Genetics*. 115 (2004):116-122

Translocations

A **translocation** occurs when a segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome. Translocations can be benign or have devastating effects depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information ([link](#)).



A reciprocal translocation occurs when a segment of DNA is transferred from one chromosome to another, nonhomologous chromosome. (credit: modification of work by National Human Genome Research/USA)

Section Summary

The number, size, shape, and banding pattern of chromosomes make them easily identifiable in a karyogram and allows for the assessment of many chromosomal abnormalities. Disorders in chromosome number, or aneuploidies, are typically lethal to the embryo, although a few trisomic genotypes are viable. Because of X inactivation, aberrations in sex chromosomes typically have milder phenotypic effects. Aneuploidies also include instances in which segments of a chromosome are duplicated or deleted. Chromosome structures may also be rearranged, for example by inversion or translocation. Both of these aberrations can result in problematic phenotypic effects. Because they force chromosomes to assume unnatural topologies during meiosis, inversions and translocations are often associated with reduced fertility because of the likelihood of nondisjunction.

Art Connections

Exercise:

Problem:

[\[link\]](#) Which of the following statements about nondisjunction is true?

- a. Nondisjunction only results in gametes with $n+1$ or $n-1$ chromosomes.
- b. Nondisjunction occurring during meiosis II results in 50 percent normal gametes.
- c. Nondisjunction during meiosis I results in 50 percent normal gametes.
- d. Nondisjunction always results in four different kinds of gametes.

Solution:

[\[link\]](#) B.

Review Questions

Exercise:

Problem:

Which of the following codes describes position 12 on the long arm of chromosome 13?

- a. 13p12
- b. 13q12
- c. 12p13
- d. 12q13

Solution:

B

Exercise:

Problem:

In agriculture, polyploid crops (like coffee, strawberries, or bananas) tend to produce _____.

- a. more uniformity
- b. more variety
- c. larger yields
- d. smaller yields

Solution:

C

Exercise:

Problem:

Assume a pericentric inversion occurred in one of two homologs prior to meiosis. The other homolog remains normal. During meiosis, what structure—if any—would these homologs assume in order to pair accurately along their lengths?

- a. V formation
- b. cruciform
- c. loop
- d. pairing would not be possible

Solution:

C

Exercise:

Problem: The genotype XXY corresponds to

- a. Klinefelter syndrome
- b. Turner syndrome
- c. Triplo-X
- d. Jacob syndrome

Solution:

A

Exercise:**Problem:**

Abnormalities in the number of X chromosomes tends to have milder phenotypic effects than the same abnormalities in autosomes because of _____.

- a. deletions
- b. nonhomologous recombination
- c. synapsis
- d. X inactivation

Solution:

D

Exercise:

Problem: By definition, a pericentric inversion includes the _____.

- a. centromere
- b. chiasma
- c. telomere
- d. synapse

Solution:

A

Free Response

Exercise:

Problem:

Using diagrams, illustrate how nondisjunction can result in an aneuploid zygote.

Solution:

Exact diagram style will vary; diagram should look like [\[link\]](#).

Glossary

aneuploid

individual with an error in chromosome number; includes deletions and duplications of chromosome segments

autosome

any of the non-sex chromosomes

chromosome inversion

detachment, 180° rotation, and reinsertion of a chromosome arm

euploid

individual with the appropriate number of chromosomes for their species

karyogram

photographic image of a karyotype

karyotype

number and appearance of an individual's chromosomes; includes the size, banding patterns, and centromere position

monosomy

otherwise diploid genotype in which one chromosome is missing

nondisjunction

failure of synapsed homologs to completely separate and migrate to separate poles during the first cell division of meiosis

paracentric

inversion that occurs outside of the centromere

pericentric

inversion that involves the centromere

polyploid

individual with an incorrect number of chromosome sets

translocation

process by which one segment of a chromosome dissociates and reattaches to a different, nonhomologous chromosome

trisomy

otherwise diploid genotype in which one entire chromosome is duplicated

X inactivation

condensation of X chromosomes into Barr bodies during embryonic development in females to compensate for the double genetic dose